Connecting the dots: how to make ETS work with ARIMA

Ivan Svetunkov

ISF2022

12th July 2022

Marketing Analytics and Forecasting | Lancaster University Management School

Introduction

Introduction

In almost any forecasting course/textbook we are taught that:

- There is Exponential Smoothing (ETS) (Hyndman et al., 2008);
- There is ARIMA (Box and Jenkins, 1976);
- Some ETS models have equivalent ARIMAs (e.g. ETS(A,N,N) and ARIMA(0,1,1));
- They are two different families of models;
- They should not be compared based on in-sample measures (e.g. AIC).

SSOE

Snyder (1985) proposed a Single Source of Error state space model, encompassing ARIMA and some ES methods.

Ord et al. (1997) extended the approach to encompass all ES methods (called ETS):

$$\begin{aligned} y_t &= w(\mathbf{v}_{t-1}) + r(\mathbf{v}_{t-1})\epsilon_t \\ \mathbf{v}_t &= f(\mathbf{v}_{t-1}) + g(\mathbf{v}_{t-1})\epsilon_t \end{aligned}, \tag{1}$$

where $\mathbf{v}_t$ is the state vector, $w(\cdot)$ is the measurement, $r(\cdot)$ is the error, $f(\cdot)$ is the transition and $g(\cdot)$ is the persistence functions.

Hyndman et al. (2002) developed an automatic model selection methodology for ETS.

ARIMA in SSOE

Hyndman et al. (2008), Section 11.5 demonstrated how pure additive SSOE relates to ARIMA.

Svetunkov and Boylan (2020) implemented the model and adapted it to the short time series in supply chain.

ES vs ARIMA

Muth (1960) showed that SES is a special case of ARIMA(0,1,1).

Nerlove and Wage (1964) showed the connection between Holt's method and ARIMA(0,2,2).

Roberts (1982) proposed damped trend model and showed its connection with ARIMA(1,1,2).

Box and Jenkins (1976) showed that several ES methods could be considered special cases of ARIMA.

McKenzie (1976) showed that pure additive Holt-Winters is equivalent to SARIMA(0,1,m+1)(0,1,0)$_m$

Chatfield (1977) showed that there are no ARIMA analogues for mixed ES methods.

ES + ARIMA, the sequential approach

Gardner (1985) found that ES forecasts can be improved by introduction of AR(1) residuals.

Taylor (2003) showed that ES with AR(1) works better than without it.

De Livera (2010) developed BATS which uses ARMA residuals for exponential smoothing based model.

De Livera et al. (2011) developed TBATS based on that.

ETS + ARIMA?

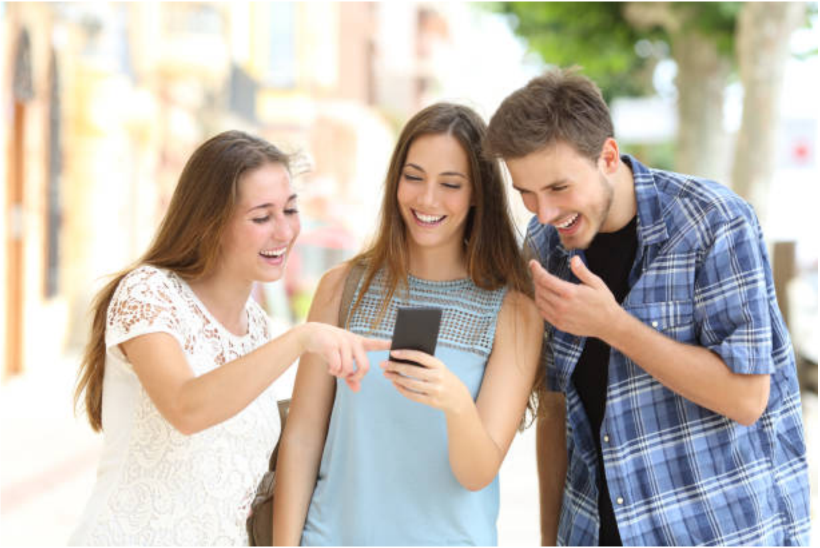
What's wrong with the sequential approach?

If ARMA is applied to ETS residuals, we introduce bias in parameters estimation.

Several research questions:

- What does ARMA residuals in ETS really mean?
- Can we merge ETS and ARIMA into ETS+ARIMA?
- What does it imply?
- What should we do with ETS(M,*,*) + ARIMA?
- How do we estimate it?

Pure additive ETS+ARIMA



Pure additive ETS

An example of a pure additive ETS(A,A,A):

$$\begin{aligned} y_t &= l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha\epsilon_t \\ b_t &= b_{t-1} + \beta\epsilon_t \\ s_t &= s_{t-m} + \gamma\epsilon_t \end{aligned}. \quad (2)$$

or

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix} + \epsilon_t \\ \begin{pmatrix} l_t \\ b_t \\ s_t \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \epsilon_t \end{aligned}. \quad (3)$$

Pure additive ETS

$$y_t = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix} + \epsilon_t$$
$$\begin{pmatrix} l_t \\ b_t \\ s_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \epsilon_t .$$

Use matrix notations:

$$\mathbf{w} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{F} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix},$$
$$\mathbf{v}_t = \begin{pmatrix} l_t \\ b_t \\ s_t \end{pmatrix}, \mathbf{v}_{t-\mathbf{l}} = \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix}, \quad \mathbf{l} = \begin{pmatrix} 1 \\ 1 \\ m \end{pmatrix}, \qquad (4)$$

Pure additive ETS

And get a pure additive ETS (Svetunkov, 2022a):

$$\begin{aligned} y_t &= \mathbf{w}'\mathbf{v}_{t-\mathbf{l}} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t-\mathbf{l}} + \mathbf{g}\epsilon_t \end{aligned}, \quad (5)$$

where $\mathbf{w}$ is called “the measurement vector”, $\mathbf{F}$ is the transition matrix, $\mathbf{g}$ is the persistence vector and $\mathbf{v}_{t-\mathbf{l}}$ is the vector of lagged components and $\mathbf{l}$ is the vector of lags.

This is a part of “ADAM” framework of Svetunkov (2022a).

Pure additive ARIMA

In a similar fashion, using (5), we could formulate ARIMA model.

Based on the idea of Hyndman et al. (2008), Section 11.5, Svetunkov (2022a) developed SSOE ARIMA.

e.g. ARIMA(0,2,2) can be formulated as:

$$\begin{aligned} y_t &= v_{1,t-1} + v_{2,t-2} + \epsilon_t \\ v_{1,t} &= 2\left(v_{1,t-1} + v_{2,t-2}\right) + \left(2 + \theta_1\right)\epsilon_t \\ v_{2,t} &= -1(v_{1,t-1} + v_{2,t-2}) + \left(-1 + \theta_2\right)\epsilon_t \end{aligned} , \quad (6)$$

Pure additive ETS+ARIMA

We can merge the two models in one:

$$\begin{aligned} y_t &= \mathbf{w}'_E \mathbf{v}_{E,t-\mathbf{l}_E} + \mathbf{w}'_A \mathbf{v}_{A,t-\mathbf{l}_A} + \epsilon_t \\ \mathbf{v}_{E,t} &= \mathbf{F}_E \mathbf{v}_{E,t-\mathbf{l}_E} + \mathbf{g}_E \epsilon_t \\ \mathbf{v}_{A,t} &= \mathbf{F}_A \mathbf{v}_{A,t-\mathbf{l}_A} + \mathbf{g}_A \epsilon_t \end{aligned}, \qquad (7)$$

where the subscript E stands for ETS and A stands for ARIMA components.

Or we could just expand the matrices $\mathbf{w}$, $\mathbf{F}$, $\mathbf{g}$, $\mathbf{v}_t$ and $\mathbf{l}$ to include first ETS and then ARIMA components.

Identifiability of ETS+ARIMA

But there is a problem with ETS+ARIMA – **identifiability**.

Consider an ETS(A,N,N)+ARIMA(0,1,1) model.

We know that SES underlies both of them. The model is formulated as:

$$\begin{aligned} y_t &= l_{t-1} + v_{1,t-1} + \epsilon_t \\ l_t &= l_{t-1} + \alpha \epsilon_t \\ v_{1,t} &= v_{1,t-1} + (1+\theta_1)\epsilon_t \end{aligned}, \quad (8)$$

The last two equations duplicate each other – they have exactly the same mechanism of update of states.

We know from Muth (1960) that the two models are equivalent when $\alpha = 1 + \theta_1$.

Identifiability of ETS+ARIMA

This means that we should avoid some combinations of models

We can also develop the following set of rules:

- For ETS(A,N,N)+ARIMA(0,1,q):
  - use ARIMA(0,1,q) in case of $q > 1$,
  - use ETS(A,N,N) in case of $q \leq 1$,
- For ETS(A,A,N)+ARIMA(0,2,q):
  - use ARIMA(0,2,q) in case of $q > 2$,
  - use ETS(A,A,N) in case of $q \leq 2$,
- For ETS(A,Ad,N)+ARIMA(p,1,q):
  - use ARIMA(p,1,q), when either $p > 1$ or $q > 2$,
  - use ETS(A,Ad,N), when $p \leq 1$ and $q \leq 2$

Other ETS+ARIMA models

Pure multiplicative ETS and logARIMA

Pure multiplicative ETS (Svetunkov, 2022a):

$$\begin{aligned} \log y_t =& \mathbf{w}' \log(\mathbf{v}_{t-\mathbf{l}}) + \log(1+\epsilon_t) \\ \log \mathbf{v}_t =& \mathbf{F} \log \mathbf{v}_{t-\mathbf{l}} + \log(\mathbf{1}_k + \mathbf{g}\epsilon_t) \end{aligned} . \quad (9)$$

We could formulate multiplicative ARIMA (or logARIMA) as (Svetunkov, 2022a):

$$\begin{aligned} & y_t = \exp\left(\sum_{j=1}^K \log v_{j,t-j} + \log(1+\epsilon_t)\right) \\ & \log v_{i,t} = \eta_i \sum_{j=1}^K \log v_{j,t-j} + (\eta_i+\theta_i)\log(1+\epsilon_t) \text{ for each } i = \{1,2,\dots,K\} \end{aligned} . \quad (10)$$

Pure multiplicative ETS+ARIMA

Merging the two together in one model:

$$
\begin{aligned}
y_t &= \exp\left(\mathbf{w}'_E \log \mathbf{v}_{E,t-\mathbf{l}_E} + \mathbf{w}'_A \log \mathbf{v}_{A,t-\mathbf{l}_A} + \log(1+\epsilon_t)\right) \\
\log \mathbf{v}_{E,t} &= \mathbf{F}_E \log \mathbf{v}_{E,t-\mathbf{l}_E} + \log(\mathbf{1}_k + \mathbf{g}_E \epsilon_t) \\
\log \mathbf{v}_{A,t} &= \mathbf{F}_A \log \mathbf{v}_{A,t-\mathbf{l}_A} + \mathbf{g}_A \log(1+\epsilon_t)
\end{aligned}
, \quad (11)
$$

with the same subscripts E and A for ETS and ARIMA respectively.

Identifiability of ETS+ARIMA

Given that there's no direct relation between ETS(M,Y,Y) and logARIMA, we do not have the same issue as before.

But some combinations can be potentially unidentifiable:

- ETS(M,N,N)+logARIMA(0,1,1),
- ETS(M,M,N)+logARIMA(0,2,2),
- ETS(M,Md,N)+logARIMA(1,1,2).

Other ETS+ARIMA combinations

Mixed ETS models can be merged with ARIMA as well.

If the error term is additive, add the additive ARIMA components.

If the error term is multiplicative, add the logARIMA components.

Mixing additive ETS with multiplicative ARIMA or vice versa does not make sense.

Model application

ETS+ARIMA estimation

We can estimate parameters of both simultaneously.

e.g. we could use likelihood, assuming a distribution for ϵ_t.

Initialisation of the ETS part is simple.

Initialisation of the ARIMA in SSOE is difficult.

Optimisation of states can be done for simple models.

Backcasting can be used for more complicated ones.

`adam()` from `smooth` (Svetunkov, 2022b) in R already does that.

ETS+ARIMA components selection

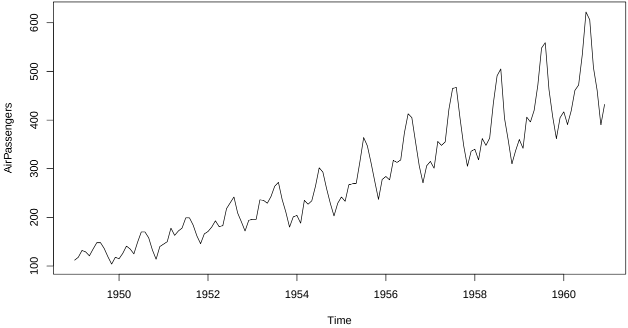
ARIMA is an addition to ETS in our framework.

This means that we can do model selection in following steps (Section 15.2 of Svetunkov, 2022a):

1. Select ETS components,
2. Extract AIC of the best ETS model (1),
3. Extract residuals,
4. Calculate ACF and PACF based on (3),
5. Try suitable ARMA model based on (4),
6. If AIC is lower than in (2), go to (3). Otherwise go to (7),
7. Reestimate the final ETS+ARIMA model.

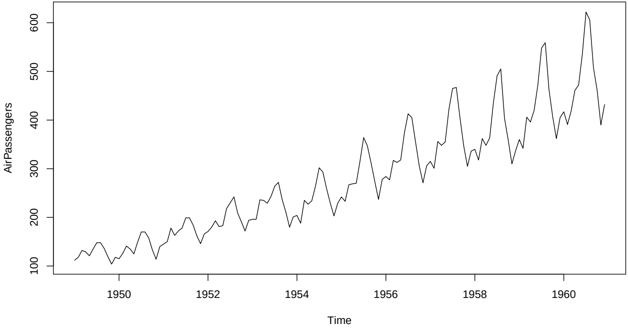
Demonstration

Consider `AirPassengers` data from R (R Core Team, 2022).



Demonstration

Consider `AirPassengers` data from R (R Core Team, 2022).



Demonstration

Start by fitting the best ETS model based on AICc:

```
model1 <- adam(AirPassengers, model="ZXZ", h=12,
holdout=TRUE, initial="back", distribution="dnorm")
```

```
AICc(model1)
```

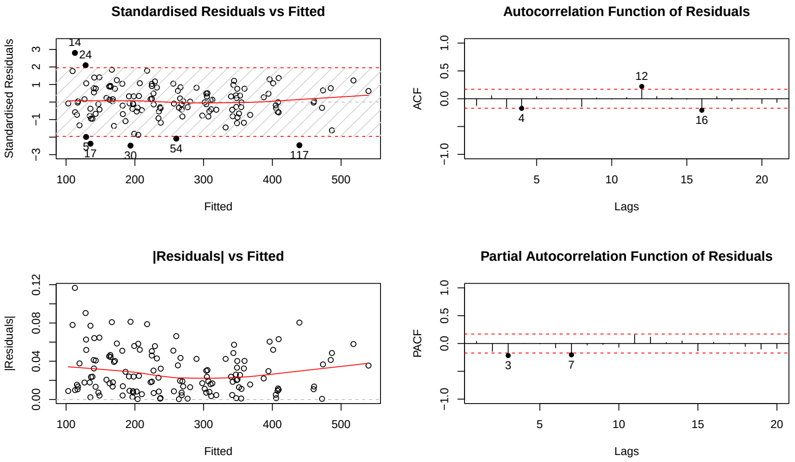
```
> 955.9526
```

Produce diagnostics plots:

```
par(mfcol=c(2,2))
```

```
plot(test,c(2,4,10,11))
```

Demonstration



Demonstration

Add AR(3):

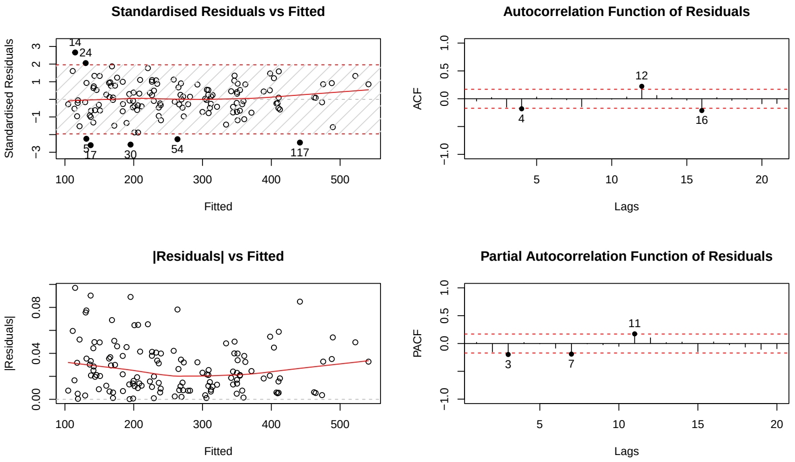
```
model2 <- adam(AirPassengers, model="MAM", h=12,
holdout=TRUE, initial="back", distribution="dnorm",
orders=list(ar=3))
```

```
AICc(model2)
```

```
> 949.7981
```

Produce diagnostics plots again...

Demonstration



Demonstration

Add Seasonal AR(1):

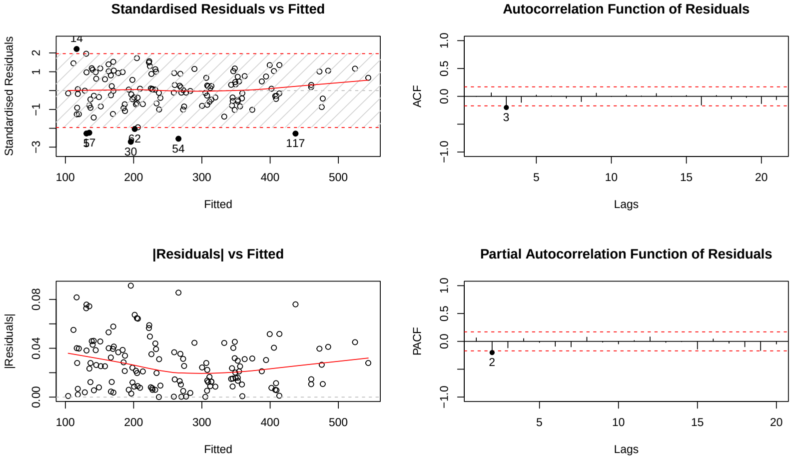
```
model3 <- adam(AirPassengers, model="MAM", h=12,
holdout=TRUE, initial="back", distribution="dnorm",
orders=list(ar=c(3,1)))

AICc(model3)

> 946.8309
```

Produce diagnostics plots again...

Demonstration



Demonstration

Try out AR(2) instead of AR(3):

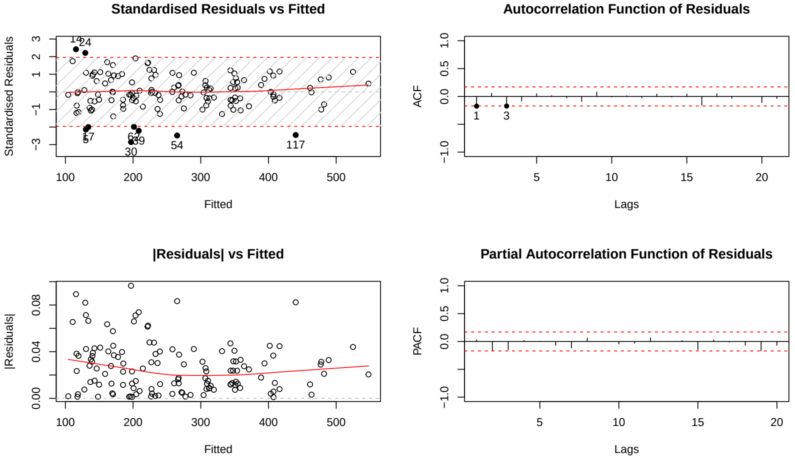
```
model4 <- adam(AirPassengers, model="MAM", h=12,
holdout=TRUE, initial="back", distribution="dnorm",
orders=list(ar=c(2,1)))
```

```
AICc(model4)
```

```
> 946.4251
```

Produce diagnostics plots again...

Demonstration



Demonstration

Final iteration, introducing MA(1)

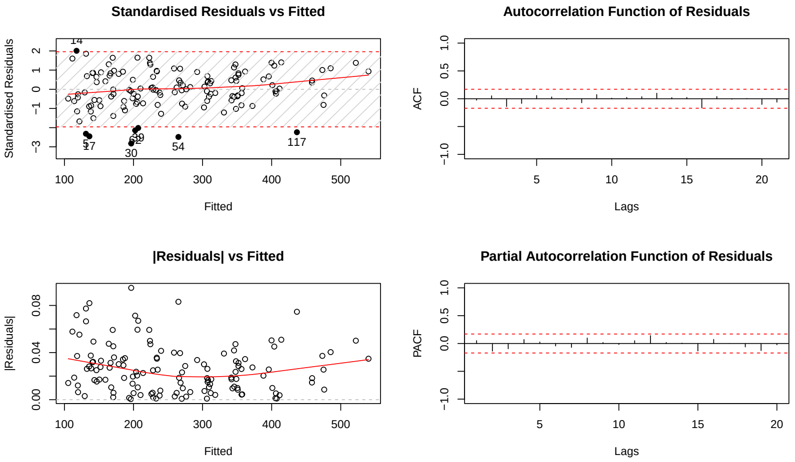
```
model5 <- adam(AirPassengers, model="MAM", h=12,
holdout=TRUE, initial="back", distribution="dnorm",
orders=list(ar=c(2,1),ma=1))

AICc(model5)

> 942.3893
```

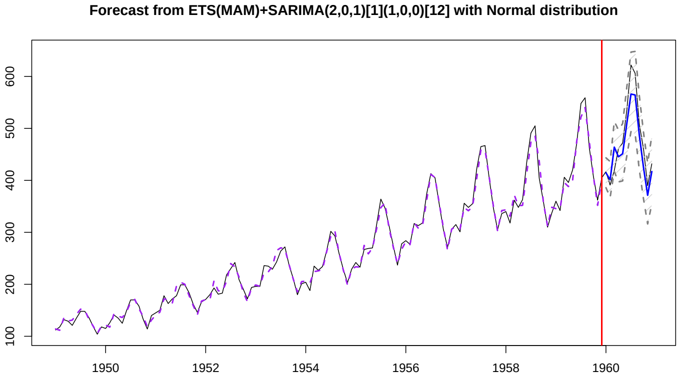
And final diagnostics...

Demonstration



Demonstration

And a forecast from this model



Demonstration

An alternative – automatic model selection:

(we'll try to select the pure model)

```
model6 <- adam(AirPassengers, model="PPP",
orders=list(ar=c(3,3),i=c(2,1),ma=c(3,3),select=TRUE),
h=12, holdout=TRUE, initial="back",
distribution="dnorm")
```

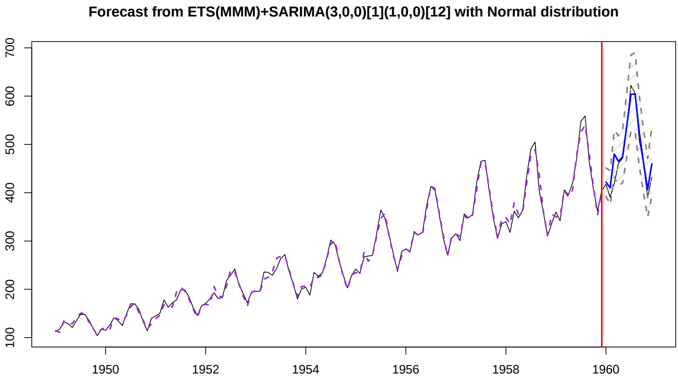
```
AICc(model6)
```

```
> 938.9278
```

Slightly better than the manual one...

Demonstration

And slightly more accurate:



Conclusions

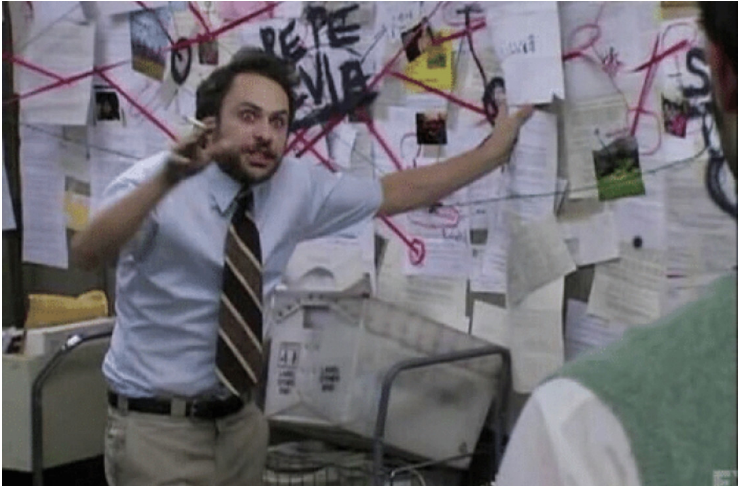


Figure: Have we connected all the dots?

Conclusions

- ETS and ARIMA can coexist in one model;
- Both can be formulated in the same SSOE framework;
- There are some non-viable combinations of the two that should be avoided;
- Estimation of the model is relatively simple;
- The model can be constructed manually via diagnostics;
- Or automatically selected based on IC minimisation.

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

https://forecasting.svetunkov.ru

https://openforecast.org/adam

twitter: @iSvetunkov

Marketing Analytics and Forecasting

Lancaster University Management School

Identifiability of ETS+ARIMA

If we estimate ETS(A,N,N)+ARIMA(0,1,1) then we duplicate the state, splitting it into two parts with some arbitrary weights.

This becomes apparent if we insert the transition equations in the measurement one:

$$\begin{aligned} y_t = & l_{t-2} + \alpha \epsilon_{t-1} + v_{1,t-2} + (1+\theta_1)\epsilon_{t-1} + \epsilon_t = \\ & l_{t-2} + v_{1,t-2} + (1+\theta_1+\alpha)\epsilon_{t-1} + \epsilon_t \end{aligned}, \quad (12)$$

There is an infinite combination of α and θ and initial values of l_0 and $v_{1,0}$ that will result in exactly the same model fit.

Thus the ETS(A,N,N)+ARIMA(0,1,1) is **unidentifiable**.

References I

Box, G., Jenkins, G., 1976. Time series analysis: forecasting and control. Holden-day, Oakland, California.

Chatfield, C., 1977. Some Recent Developments in Time-Series Analysis. Journal of the Royal Statistical Society. Series A (General) 140 (4), 492.
URL https://www.jstor.org/stable/2345281?origin=crossref

De Livera, A. M., 2010. Exponentially weighted methods for multiple seasonal time series. International Journal of Forecasting 26 (4), 655–657.
URL http://dx.doi.org/10.1016/j.ijforecast.2010.05.010

References II

De Livera, A. M., Hyndman, R. J., Snyder, R. D., 2011. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. Journal of the American Statistical Association 106 (496), 1513–1527.

Gardner, E. S., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4 (1), 1–28. URL `http://doi.wiley.com/10.1002/for.3980040103`

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing. Springer Berlin Heidelberg.

References III

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., jul 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.
URL http://linkinghub.elsevier.com/retrieve/pii/S0169207001001108

McKenzie, E., mar 1976. A Comparison of Some Standard Seasonal Forecasting Systems. The Statistician 25 (1), 3.
URL https://www.jstor.org/stable/10.2307/2988127?origin=crossref

Muth, J. F., 1960. Optimal Properties of Exponentially Weighted Forecasts. Journal of the American Statistical Association 55 (1), 299–306.

References IV

Nerlove, M., Wage, S., 1964. On the Optimality of Adaptive Forecasting. Management Science 10 (2), 207–224.

Ord, J. K., Koehler, A. B., Snyder, R. D., dec 1997. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. Journal of the American Statistical Association 92 (440), 1621–1629.
URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1997.10473684

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

References V

Roberts, S. A., 1982. A General Class of Holt-Winters Type Forecasting Models. Management Science 28 (7), 808–820.

Snyder, R. D., 1985. Recursive Estimation of Dynamic Linear Models. Journal of the Royal Statistical Society, Series B (Methodological) 47 (2), 272–276.

Svetunkov, I., 2022a. Forecasting and analytics with adam. Monograph. OpenForecast, (version: 2022-06-10). URL `https://openforecast.org/adam/`

Svetunkov, I., 2022b. smooth: Forecasting Using State Space Models. R package version 3.2.0. URL `https://github.com/config-i1/smooth`

References VI

Svetunkov, I., Boylan, J. E., feb 2020. State-space ARIMA for supply-chain forecasting. International Journal of Production Research 58 (3), 818–827.
URL https://www.tandfonline.com/doi/full/10.1080/00207543.2019.1600764

Taylor, J. W., 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of the Operational Research Society 54 (8), 799–805.