Exponential smoothing and non-negative data

Muhammad Akram
Rob J Hyndman
J Keith Ord

Business & Economic Forecasting Unit
MONASH University

Outline

Outline

Problem

- Most forecasting methods in business are based on exponential smoothing.

Problem

- Most forecasting methods in business are based on exponential smoothing.
- Most time series in business are inherently non-negative.

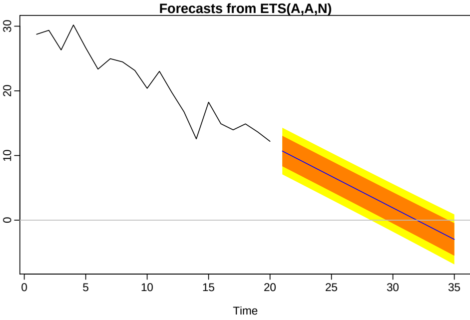
Problem

- Most forecasting methods in business are based on exponential smoothing.
- Most time series in business are inherently non-negative.
- But exponential smoothing models for non-negative data don't always work!

Problem

- Most forecasting methods in business are based on exponential smoothing.
- Most time series in business are inherently non-negative.
- But exponential smoothing models for non-negative data don't always work!

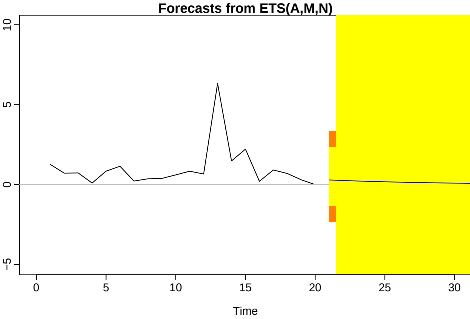
1. They can produce negative forecasts



Problem

- Most forecasting methods in business are based on exponential smoothing.
- Most time series in business are inherently non-negative.
- But exponential smoothing models for non-negative data don't always work!

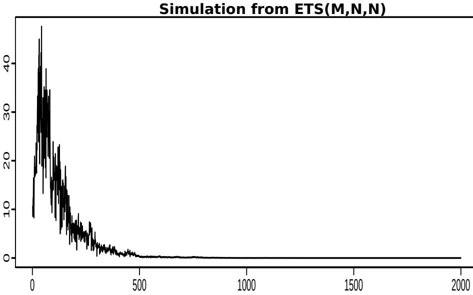
1. They can produce negative forecasts
2. They can produce infinite forecast variance



Problem

- Most forecasting methods in business are based on exponential smoothing.
- Most time series in business are inherently non-negative.
- But exponential smoothing models for non-negative data don't always work!

1. They can produce negative forecasts
2. They can produce infinite forecast variance
3. They can converge almost surely to zero.



Taxonomy of models

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

Taxonomy of models

		Seasonal Component		
Trend Component		N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)

Taxonomy of models

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)

Exponen**T**ial **S**moothing

Taxonomy of models

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	**N,N**	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)
Exponen**T**ial **S**moothing

ETS(A,N,N): Simple exponential smoothing with additive errors

Taxonomy of models

Trend Component		**Seasonal Component** N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	**A,N**	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)

Exponen**T**ial **S**moothing

ETS(A,A,N): Holt's linear method with additive errors

Taxonomy of models

		Seasonal Component		
Trend Component		N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	**A,A**	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)
Exponen**T**ial **S**moothing

ETS(A,A,A): Additive Holt-Winters' method with additive errors

Taxonomy of models

		Seasonal Component		
Trend Component		N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	**A,M**
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)
Exponen**T**ial **S**moothing

ETS(M,A,M): Multiplicative Holt-Winters' method with multiplicative errors

Taxonomy of models

Trend Component		N (None)	A (Additive)	M (Multiplicative)
		Seasonal Component		
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	**A_d,N**	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)
Exponen**T**ial **S**moothing

ETS(A,A_d,N): Damped trend method with additive errors

Taxonomy of models

Trend Component		Seasonal Component: N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation **ETS**(*Error,Trend,Seasonal*)

Exponen**T**ial **S**moothing

There are 30 separate models in the ETS framework

Innovations state space model

No trend or seasonality and multiplicative errors

Example: ETS(M,N,N)

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

$0 \leq \alpha \leq 1$

ε_t is white noise with mean zero.

Innovations state space model

No trend or seasonality and multiplicative errors

Example: ETS(M,N,N)

$$y_t = \ell_{t-1}(1+\varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$$

$0 \le \alpha \le 1$

ε_t is white noise with mean zero.

Innovations state space model

No trend or seasonality and multiplicative errors

Example: ETS(M,N,N)

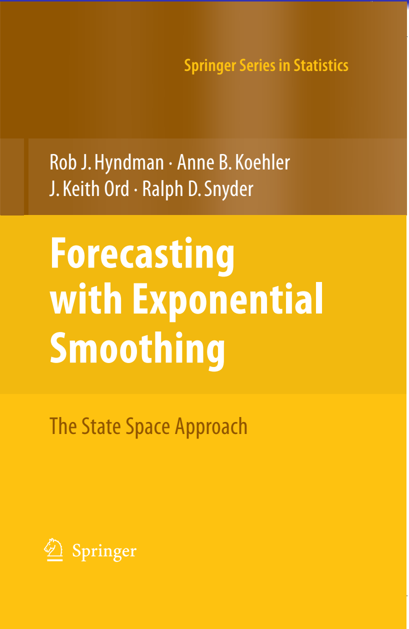
$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$$

$0 \le \alpha \le 1$

ε_t is white noise with mean zero.

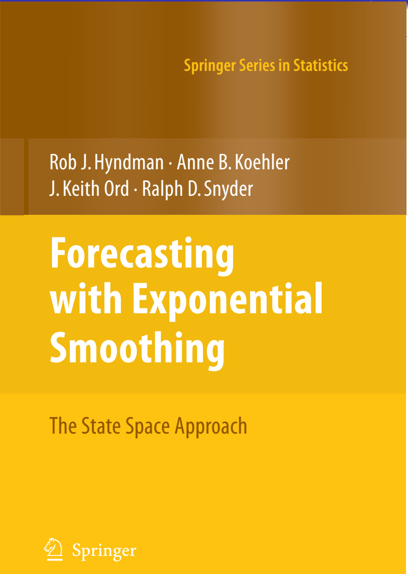
All exponential smoothing models can be written using analogous state space equations.

New book!



- State space modeling framework
- Prediction intervals
- Model selection
- Maximum likelihood estimation
- All the important research results in one place with consistent notation
- Many new results
- 375 pages but only US$39.95.

New book!

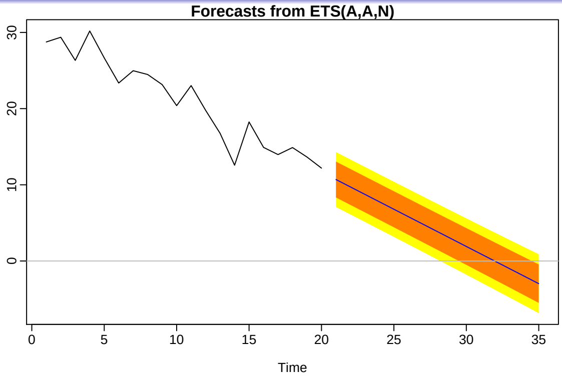


- State space modeling framework
- Prediction intervals
- Model selection
- Maximum likelihood estimation
- All the important research results in one place with consistent notation
- Many new results
- 375 pages but only US$39.95.

www.exponentialsmoothing.net

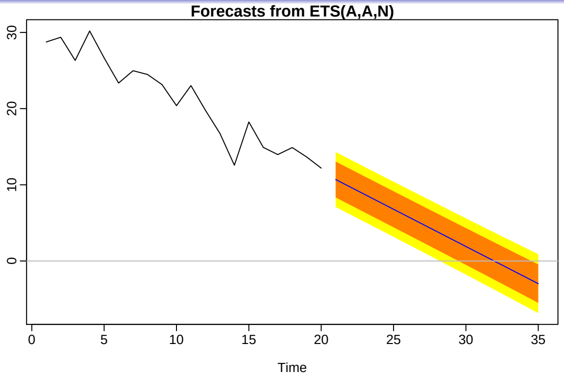
Outline

Negative forecasts



- Could solve by taking logs or some other Box-Cox transformation. However, this limits models to be additive in the transformed space.

Negative forecasts



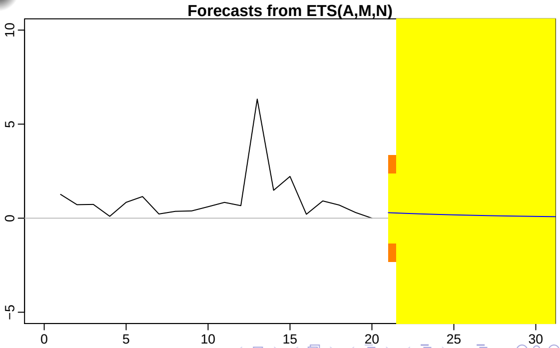
- Could solve by taking logs or some other Box-Cox transformation. However, this limits models to be additive in the transformed space.
- Could solve by only using multiplicative models. But these can have other problems.

Infinite forecast variances

ETS(A,M,N) model

$$y_t = \ell_{t-1} b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$$
$$b_t = b_{t-1} + \beta \varepsilon_t / \ell_{t-1}.$$

- $\ell_0 = 0.1$
- $b_0 = 1$
- $\alpha = 0.1$
- $\beta = 0.05$
- $\sigma = 1$

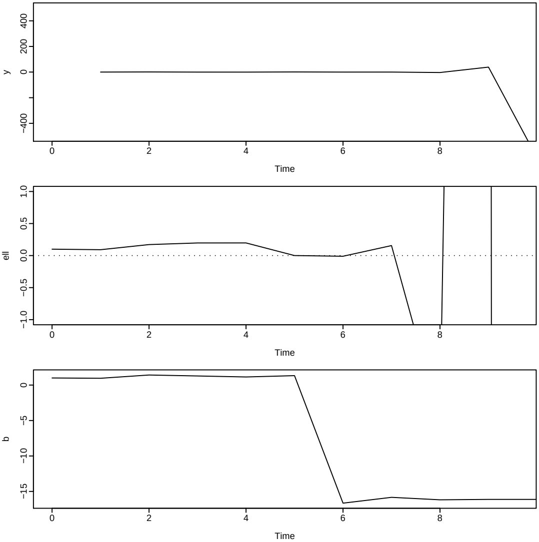


Infinite forecast variances

ETS(A,M,N) model

$$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}.$$

- $\ell_0 = 0.1$
- $b_0 = 1$
- $\alpha = 0.1$
- $\beta = 0.05$
- $\sigma = 1$



Infinite forecast variances

Suppose ε_t has positive density at 0

For ETS models (A,M,N), (A,M,A), (A,M_d,N), (A,M_d,A), (A,M,M), (A,M_d,M), (M,M,A) and (M,M_d,A):

- $\text{V}(y_{n+h} \mid \boldsymbol{x}_n) = \infty$ for $h \geq 3$;
- $\text{E}(y_{n+h} \mid \boldsymbol{x}_n)$ is undefined for $h \geq 3$.

Infinite forecast variances

Suppose ε_t has positive density at 0

For ETS models (A,M,N), (A,M,A), (A,M$_\text{d}$,N), (A,M$_\text{d}$,A), (A,M,M), (A,M$_\text{d}$,M), (M,M,A) and (M,M$_\text{d}$,A):

- $\text{V}(y_{n+h} \mid \boldsymbol{x}_n) = \infty$ for $h \geq 3$;
- $\text{E}(y_{n+h} \mid \boldsymbol{x}_n)$ is undefined for $h \geq 3$.

For ETS models (A,N,M), (A,A,M) and (A,A$_\text{d}$,M):

- $\text{V}(y_{n+h} \mid \boldsymbol{x}_n) = \infty$ for $h \geq m+2$;
- $\text{E}(y_{n+h} \mid \boldsymbol{x}_n)$ is undefined for $h \geq m+2$.

Infinite forecast variances

Suppose ε_t has positive density at 0

For ETS models (A,M,N), (A,M,A), (A,M_d,N), (A,M_d,A), (A,M,M), (A,M_d,M), (M,M,A) and (M,M_d,A):

- $\text{V}(y_{n+h} \mid \boldsymbol{x}_n) = \infty$ for $h \geq 3$;
- $\text{E}(y_{n+h} \mid \boldsymbol{x}_n)$ is undefined for $h \geq 3$.

For ETS models (A,N,M), (A,A,M) and (A,A_d,M):

- $\text{V}(y_{n+h} \mid \boldsymbol{x}_n) = \infty$ for $h \geq m + 2$;
- $\text{E}(y_{n+h} \mid \boldsymbol{x}_n)$ is undefined for $h \geq m + 2$.

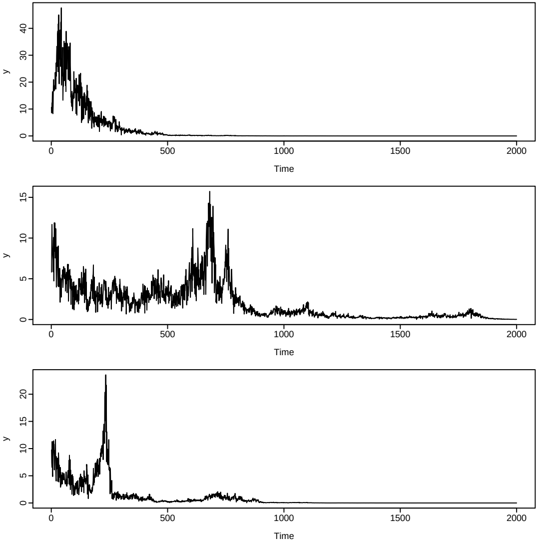
➥ **These problems occur regardless of the sample space of $\{y_t\}$.**

Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1+\varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$$

- $\ell_0 = 10$
- $\alpha = 0.3$
- $\sigma = 0.3$ with truncated Gaussian errors



Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t),$$

- $0 < \alpha \leq 1$

Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t),$$

- $0 < \alpha \leq 1$
- $\delta_t = 1 + \varepsilon_t$ has mean 1 and variance σ^2.

Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t),$$

- $0 < \alpha \leq 1$
- $\delta_t = 1 + \varepsilon_t$ has mean 1 and variance σ^2.
- δ_t are iid with positive distribution such as truncated normal, lognormal, gamma, etc.

Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t),$$

- $0 < \alpha \leq 1$
- $\delta_t = 1 + \varepsilon_t$ has mean 1 and variance σ^2.
- δ_t are iid with positive distribution such as truncated normal, lognormal, gamma, etc.

Convergence to zero

ETS(M,N,N) model

$$y_t = \ell_{t-1}(1+\varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t),$$

- $0 < \alpha \le 1$
- $\delta_t = 1 + \varepsilon_t$ has mean 1 and variance σ^2.
- δ_t are iid with positive distribution such as truncated normal, lognormal, gamma, etc.

$$\ell_t = \ell_0(1+\alpha\varepsilon_1)(1+\alpha\varepsilon_2)\cdots(1+\alpha\varepsilon_t) = \ell_0 U_t,$$

where $U_t = U_{t-1}(1+\alpha\varepsilon_t)$ and $U_0 = 1$. Thus U_t is a non-negative product martingale.

Convergence to zero

- Kakutani's Theorem says that ℓ_t will converge to 0 almost surely if ε_t has mean zero and is non-degenerate.

Convergence to zero

- Kakutani's Theorem says that ℓ_t will converge to 0 almost surely if ε_t has mean zero and is non-degenerate.
- Consequently, all sample paths for y_t converge to 0 almost surely.

Convergence to zero

- Kakutani's Theorem says that ℓ_t will converge to 0 almost surely if ε_t has mean zero and is non-degenerate.
- Consequently, all sample paths for y_t converge to 0 almost surely.
- Similar results follow for all purely multiplicative models: (M,N,N), (M,N,M), (M,M,N), (M,M,M), (M,M_d,N) and (M,M_d,M).

Four model classes

Class M: **Purely multiplicative models:** (M,N,N), (M,N,M), (M,M,N), (M,M,M), (M,M_d,N) and (M,M_d,M).

Class A: **Purely additive models:** (A,N,N), (A,N,A), (A,A,N), (A,A,A), (A,A_d,N) and (A,A_d,A).

Class X: **Mixed models:** (A,M,$*$), (A,M_d,$*$), (A,$*$,M), (M,M,A), (M,M_d,A). (11 models)

Class Y: **Mixed models:** (M,A,$*$), (M,A_d,$*$) or (M,N,A). (7 models)

Four model classes

Class M: **Purely multiplicative models:** (M,N,N), (M,N,M), (M,M,N), (M,M,M), (M,M_d,N) and (M,M_d,M).

Class A: **Purely additive models:** (A,N,N), (A,N,A), (A,A,N), (A,A,A), (A,A_d,N) and (A,A_d,A).

Class X: **Mixed models:** (A,M,$*$), (A,M_d,$*$), (A,$*$,M), (M,M,A), (M,M_d,A). (11 models)

Class Y: **Mixed models:** (M,A,$*$), (M,A_d,$*$) or (M,N,A). (7 models)

- Only Class M can guarantee a sample space restricted to the positive half-line.

Four model classes

Class M: **Purely multiplicative models:** (M,N,N), (M,N,M), (M,M,N), (M,M,M), (M,M_d,N) and (M,M_d,M).

Class A: **Purely additive models:** (A,N,N), (A,N,A), (A,A,N), (A,A,A), (A,A_d,N) and (A,A_d,A).

Class X: **Mixed models:** (A,M,$*$), (A,M_d,$*$), (A,$*$,M), (M,M,A), (M,M_d,A). (11 models)

Class Y: **Mixed models:** (M,A,$*$), (M,A_d,$*$) or (M,N,A). (7 models)

- Only Class M can guarantee a sample space restricted to the positive half-line.
- All Class M models converge to 0 if $\text{E}(\varepsilon) = 0$

Four model classes

Class M: **Purely multiplicative models:** (M,N,N), (M,N,M), (M,M,N), (M,M,M), (M,M_d,N) and (M,M_d,M).

Class A: **Purely additive models:** (A,N,N), (A,N,A), (A,A,N), (A,A,A), (A,A_d,N) and (A,A_d,A).

Class X: **Mixed models:** (A,M,$*$), (A,M_d,$*$), (A,$*$,M), (M,M,A), (M,M_d,A). (11 models)

Class Y: **Mixed models:** (M,A,$*$), (M,A_d,$*$) or (M,N,A). (7 models)

- Only Class M can guarantee a sample space restricted to the positive half-line.
- All Class M models converge to 0 if E$(\varepsilon) = 0$
- All Class X models have infinite forecast variance for $h \geq m + 2$ where m is the seasonal period.

Outline

New models

Let $\delta_t = (1 + \varepsilon_t)$ be a positive random variable.

METS(M,N,N) model

$$y_t = \ell_{t-1}\delta_t$$
$$\ell_t = \ell_{t-1}\delta_t^{\alpha}$$

New models

Let $\delta_t = (1+\varepsilon_t)$ be a positive random variable.

METS(M,N,N) model

$$y_t = \ell_{t-1}\delta_t$$
$$\ell_t = \ell_{t-1}\delta_t^{\alpha}$$

$$\log(y_t) = \log(\ell_{t-1}) + \log(\delta_t)$$
$$\log(\ell_t) = \log(\ell_{t-1}) + \alpha\log(\delta_t)$$

New models

Let $\delta_t = (1 + \varepsilon_t)$ be a positive random variable.

METS(M,N,N) model

$$y_t = \ell_{t-1}\delta_t$$
$$\ell_t = \ell_{t-1}\delta_t^{\alpha}$$

$$\log(y_t) = \log(\ell_{t-1}) + \log(\delta_t)$$
$$\log(\ell_t) = \log(\ell_{t-1}) + \alpha \log(\delta_t)$$

Thus the log-transformed model is identical to Gaussian ETS(A,N,N) model if δ_t is logNormal with median 1.

Long term forecast behaviour

METS(M,N,N; LN) model

$$y_t = \ell_{t-1}\delta_t$$
$$\ell_t = \ell_{t-1}\delta_t^{\alpha},$$

$$\delta_t \sim \text{logN}(\mu, \omega)$$

Long term forecast behaviour

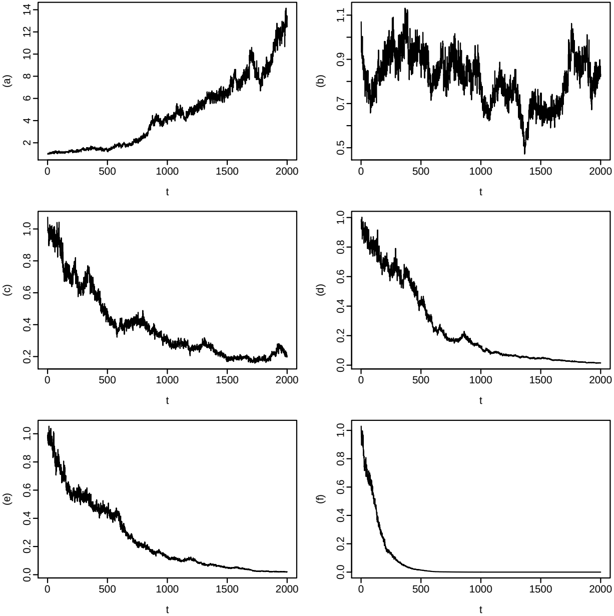
METS(M,N,N; LN) model

$$y_t = \ell_{t-1}\delta_t$$
$$\ell_t = \ell_{t-1}\delta_t^{\alpha},$$

$$\delta_t \sim \text{logN}(\mu, \omega)$$

Range	$E(\delta_t^{\alpha})$	$E(\delta_t^{\alpha/2})$	$E(y_h)$	$V(y_h)$
$\mu + \alpha\omega < 0$	< 1	< 1	Decreasing	Decreasing
$\mu + \alpha\omega = 0$	< 1	< 1	Decreasing	Finite
$-\alpha\omega < \mu < -\alpha\omega/2$	< 1	< 1	Decreasing	Increasing
$\mu + \alpha\omega/2 = 0$	$= 1$	< 1	Finite	Increasing
$-\alpha\omega/2 < \mu < -\alpha\omega/4$	> 1	< 1	Increasing	Increasing
$\mu + \alpha\omega/4 = 0$	> 1	$= 1$	Increasing	Increasing
$\mu + \alpha\omega/4 > 0$	> 1	> 1	Increasing	Increasing

Long term forecast behaviour



METS(M,N,N;LN)

$\delta_t \sim \text{logN}(\mu, \omega)$:

(a) $\mu = \alpha\omega/4$

(b) $\mu = 0$

(c) $\mu = -\alpha\omega/4$

(d) $\mu = -3\alpha\omega/8$

(e) $\mu = -\alpha\omega/2$

(f) $\mu = -3\alpha\omega/4$

Outline

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.
3. An alternative formulation has been proposed that avoids these problems.

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.
3. An alternative formulation has been proposed that avoids these problems.
4. The ergodic behaviour of the alternative formulation requires careful parameter choice.

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.
3. An alternative formulation has been proposed that avoids these problems.
4. The ergodic behaviour of the alternative formulation requires careful parameter choice.

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.
3. An alternative formulation has been proposed that avoids these problems.
4. The ergodic behaviour of the alternative formulation requires careful parameter choice.

Paper: **www.robhyndman.info**

Conclusions

1. The standard exponential smoothing state space models all have theoretical problems when constrained to the positive half line.
2. This is not a concern for most short-term forecasting.
3. An alternative formulation has been proposed that avoids these problems.
4. The ergodic behaviour of the alternative formulation requires careful parameter choice.

Paper: **www.robhyndman.info**

Book: **www.exponentialsmoothing.net**