Chapter 7 Exponential Smoothing

7.1 Simple exponential smoothing

- trend나 seasonal pattern에 적용하는 것이 적당

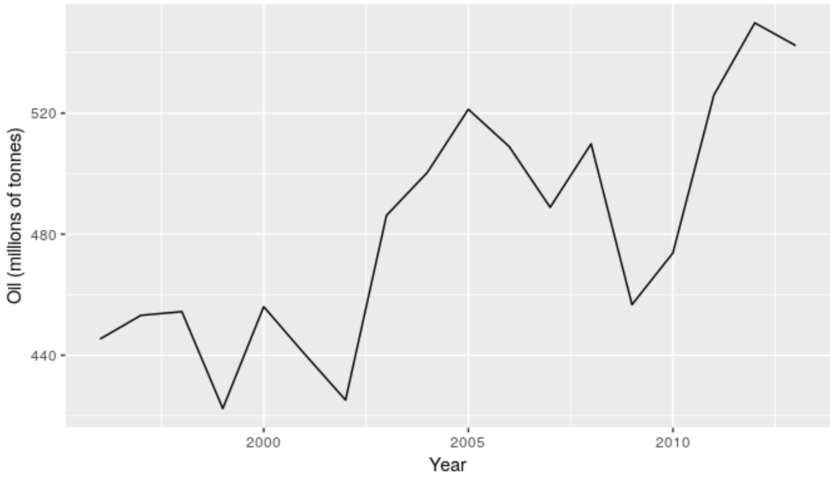


Figure 7.1: Oil production in Saudi Arabia from 1980 to 2013.

- 과거일수록 더 가중치가 낮아지는 가중합

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots, \qquad (7.1)$$

- 수학적으로는 아래처럼, 이전 step의 forcast를 현재 관측값과 가중합하는 것과 동치

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

Component form

- only component included is the level(smoothed value), l(t)
  - h = 1 이면 fitted value
  - t = T 이면 flat forcast

$$\begin{aligned} \text{Forecast equation} \qquad \hat{y}_{t+h|t} &= \ell_t \\ \text{Smoothing equation} \qquad \ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1}, \end{aligned}$$

최적화

- 위의 component form을 적용하려면, alpha와 l(0)가 parameter
- 예를 들어 SSE를 최소화하는 최적 파라미터로서 구한다.

$$\text{SSE} = \sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^{T} e_t^2. \qquad (7.2)$$

예제

- 오일 예제에서 SSE 최소화를 만족시키는 최적값은 alpha=0.83, l(0)=446.6
- 이를 이용해서 점화식에 대입하면
  - t < T 까지는 fitted value
  - t > T 부터는 flat forcast

simple exponential smoothing.

Year	Time	Observation	Level	Forecast
	t	y_t	ℓ_t	$\hat{y}_{t+1\|t}$
1995	0		446.59	
1996	1	445.36	445.57	446.59
1997	2	453.20	451.93	445.57
1998	3	454.41	454.00	451.93
1999	4	422.38	427.63	454.00
2000	5	456.04	451.32	427.63
2001	6	440.39	442.20	451.32
2002	7	425.19	428.02	442.20
2003	8	486.21	476.54	428.02
2004	9	500.43	496.46	476.54
2005	10	521.28	517.15	496.46
2006	11	508.95	510.31	517.15
2007	12	488.89	492.45	510.31
2008	13	509.87	506.98	492.45
2009	14	456.72	465.07	506.98
2010	15	473.82	472.36	465.07
2011	16	525.95	517.05	472.36
2012	17	549.83	544.39	517.05
2013	18	542.34	542.68	544.39
	h			$\hat{y}_{T+h\|T}$
2014	1			542.68
2015	2			542.68
2016	3			542.68
2017	4			542.68
2018	5			542.68

The black line in Figure 7.2 is a plot of the data, which shows a changing level over time.

- alpha가 크면, 현재 관측값이 더 중요해져서, 더 sharp해지고
- alpha가 작으면, 과거 level값이 중요해져서, 더 smooth해진다.

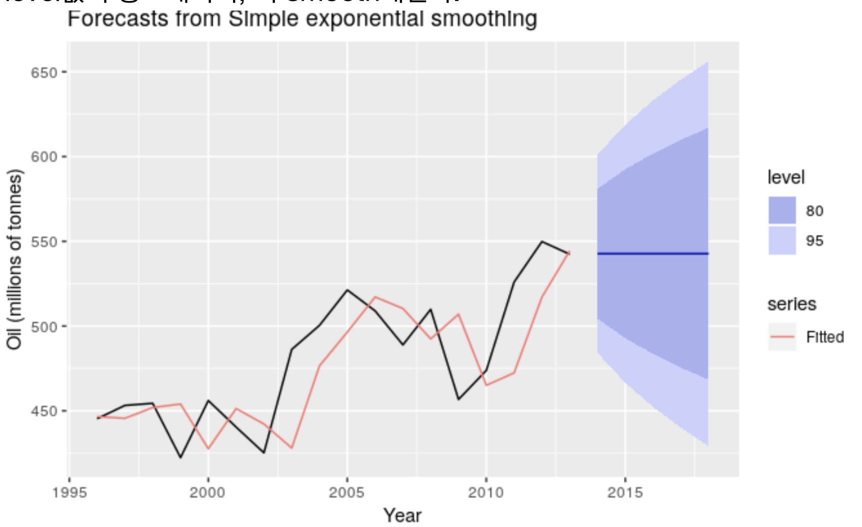


Figure 7.2: Simple exponential smoothing applied to oil production in Saudi Arabia (1996–2013).

7.2 Trend methods

Holt's linear trend method (1957)

- a forecast equation and two smoothing equations (one for the level and one for the trend)

- l(t) 는 현재 관측값인 y(t)와 one-step training forcast끼리의 가중합
- b(t) 는 현재의 slope 관측값인 l(t) - l(t-1)과 이전 step의 slope 값 끼리의 가중합

$$\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1},
\end{aligned}$$

더 이상 flat forcast가 아니다. trendy한 forcast이다.

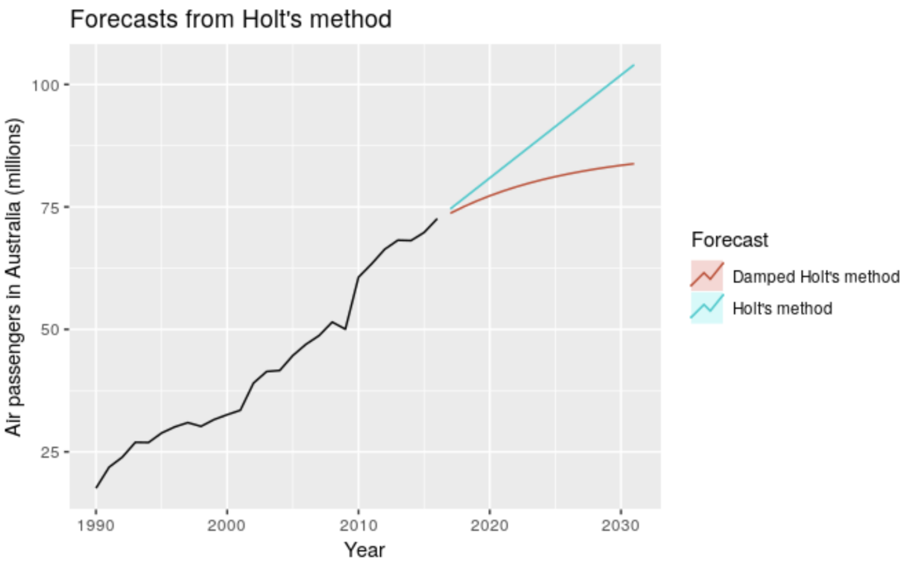
2014	25	68.12	68.49	2.102	70.31
2015	26	69.78	69.92	2.102	70.60
2016	27	72.60	72.50	2.102	72.02
	h				$\hat{y}_{t+h\|t}$
	1				74.60
	2				76.70
	3				78.80
	4				80.91
	5				83.01

Damped trend methods (1985)

- holt의 방법은 무한히 상승(하강)하는 forcast라는 말도 안되는 결과가 나온다. (h가 크다면)
- 이를 꺽는(dampen) 파리머터를 도입하자.
- pi = 1이면 holt랑 동일

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}.
\end{aligned}$$

short horizon에서는 trended, longer horizon에서는 상수값으로 수렴



7.3 Holt-Winters' seasonal method (1960)

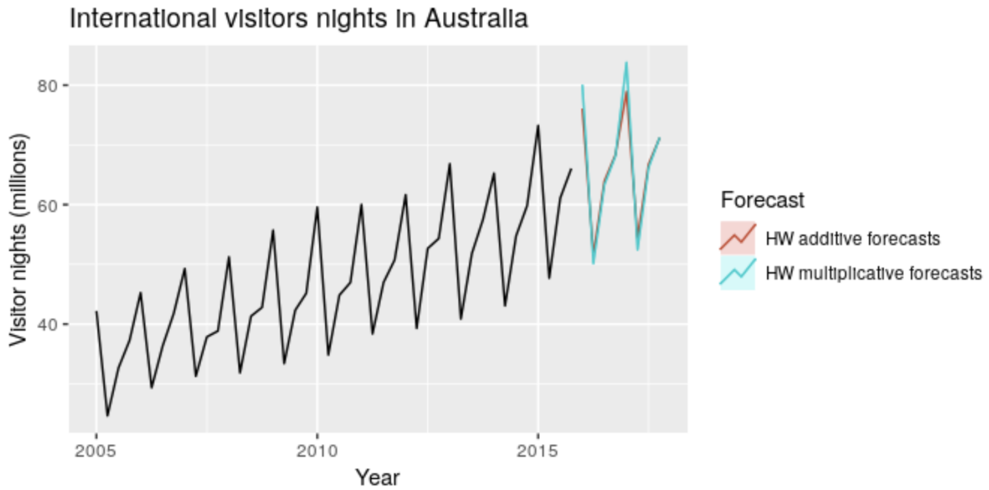
- one forecast equation and three smoothing equations (level, trend, seasonal)

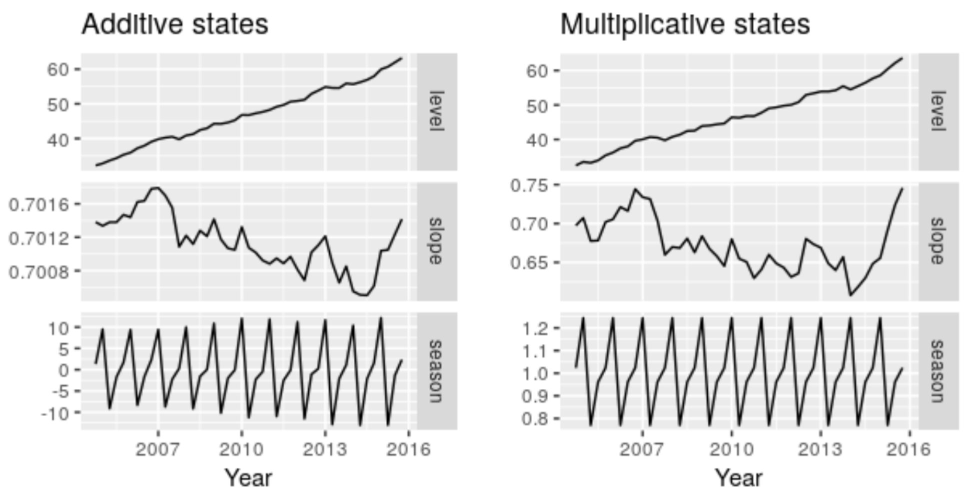
Holt-Winters' additive method

- seasonal variations are roughly constant through the series
- The level equation shows a weighted average between the seasonally adjusted observation and the non-seasonal forecast for time t
- The seasonal equation shows a weighted average between the current seasonal index, and the seasonal index of the same season last year

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\ \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}, \end{aligned}$$

예제





Holt-Winters' damped method

- Holt-Winters' multiplicative method + damping

7.4 A taxonomy of exponential smoothing methods (1969)

- Trend와 seasonal의 상황에 따라 3x3 = 9가지 경우의 수

Table 7.5: A two-way classification of exponential smoothing methods.

Trend Component	**Seasonal Component**		
	N	**A**	**M**
	(None)	**(Additive)**	**(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)

Some of these methods we have already seen using other names:

Short hand	Method
(N,N)	Simple exponential smoothing
(A,N)	Holt's linear method
(A_d,N)	Additive damped trend method
(A,A)	Additive Holt-Winters' method
(A,M)	Multiplicative Holt-Winters' method
(A_d,M)	Holt-Winters' damped method

Trend	**Seasonal** N	**Seasonal** A	**Seasonal** M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_d	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

7.5 Innovations state space models for exponential smoothing

- 지금까지는 point forcast이고, prediction interval은 없었다.
- 이제 통계적 분석을 해보자
- ETS(,,) = (Error, Trend, Seasonal)
- Error는 Additive or Multiplicative

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

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+1|t} = \ell_t \\ \text{Smoothing equation} \quad & \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}, \end{aligned}$$

If we re-arrange the smoothing equation for the level, we get the "error correction" form:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

where $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ is the residual at time t.

Then the equations of the model can be written as

$$y_t = \ell_{t-1} + \varepsilon_t \tag{7.3}$$
$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t. \tag{7.4}$$

- (7.3) measurment equation은 관측인 내재된 구성을 보여준다.
  - 예측 가능한 부분은, level
  - 예측 불가능한 부분은, error
- (7.4) state equation
  - 시간에 따라서 state가 evolve하는 것을 의미
  - alpha값에 따라서 상태의 변화 정도가 조절
  - alpha=0 이면, level이 전혀 변화지 않는 상태
  - alpha=1 이면, pure random walk

All methods

Table 7.7: State space equations for each of the models in the ETS framework.

ADDITIVE ERROR MODELS

Trend	**Seasonal** **N**	**A**	**M**
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
$\mathbf{A_d}$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

7.6 Estimation and model selection

Estimating ETS models

- 지금까지는 Mimimization of MSE 를 통해 parameter estimation을 보여줬음
  - 또다른 방법은 ML(maximum likelihood) 이다.

- Additive error 계열 방법은 MMSE나 ML이나 결과 동일하나, Multiplicative error 계열은 두 방법의 결과가 다르다.
- 지금까지는 모든 파라미터는 0과 1사이로 제약(traditional constraint)을 두었다.
  - state-space model에서는 이를 완화한(less restrictive)한 파라미터로 볼 수 있다.
  - addimissible constraint,
  - ex) 0 < alpha < 2

Model selection

For ETS models, Akaike's Information Criterion (AIC) is defined as

$$\text{AIC} = -2\log(L) + 2k,$$

where L is the likelihood of the model and k is the total number of parameters and initial states that have been estimated (including the residual variance).

The AIC corrected for small sample bias (AIC_c) is defined as

$$\text{AIC}_\text{c} = \text{AIC} + \frac{k(k+1)}{T-k-1},$$

and the Bayesian Information Criterion (BIC) is

$$\text{BIC} = \text{AIC} + k[\log(T) - 2].$$

7.7 Forecasting with ETS models

Point forecasts are obtained from the models by iterating the equations for $t = T+1, \dots, T+h$ and setting all $\varepsilon_t = 0$ for $t > T$.

For example, for model ETS(M,A,N), $y_{T+1} = (\ell_T + b_T)(1 + \varepsilon_{T+1})$. Therefore $\hat{y}_{T+1|T} = \ell_T + b_T$. Similarly,

$$\begin{aligned} y_{T+2} &= (\ell_{T+1} + b_{T+1})(1 + \varepsilon_{T+1}) \\ &= \left[(\ell_T + b_T)(1 + \alpha\varepsilon_{T+1}) + b_T + \beta(\ell_T + b_T)\varepsilon_{T+1}\right](1 + \varepsilon_{T+1}). \end{aligned}$$

Therefore, $\hat{y}_{T+2|T} = \ell_T + 2b_T$, and so on. These forecasts are identical to the forecasts from

For most ETS models, a prediction interval can be written as

$$\hat{y}_{T+h|T} \pm c\sigma_h$$

Table 7.8: Forecast variance expressions for each additive state space model, where σ^2 is the residual variance, m is the seasonal period, and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time $T+h$).

Model	Forecast variance: σ_h^2
(A,N,N)	$\sigma_h^2 = \sigma^2\big[1 + \alpha^2(h-1)\big]$
(A,A,N)	$\sigma_h^2 = \sigma^2\Big[1 + (h-1)\big\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\big\}\Big]$
(A,A$_d$,N)	$\sigma_h^2 = \sigma^2\Big[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\big\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\big\}\Big]$
(A,N,A)	$\sigma_h^2 = \sigma^2\Big[1 + \alpha^2(h-1) + \gamma k(2\alpha+\gamma)\Big]$
(A,A,A)	$\sigma_h^2 = \sigma^2\Big[1 + (h-1)\big\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\big\} + \gamma k\big\{2\alpha + \gamma + \beta m(k+1)\big\}\Big]$
(A,A$_d$,A)	$\sigma_h^2 = \sigma^2\Big[1 + \alpha^2(h-1) + \gamma k(2\alpha+\gamma) + \frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\big\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\big\} + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\big\{k(1-\phi^m) - \phi^m(1-\phi^{mk})\big\}\Big]$

In [ ]: