

Блеск и нищета процедур сглаживания



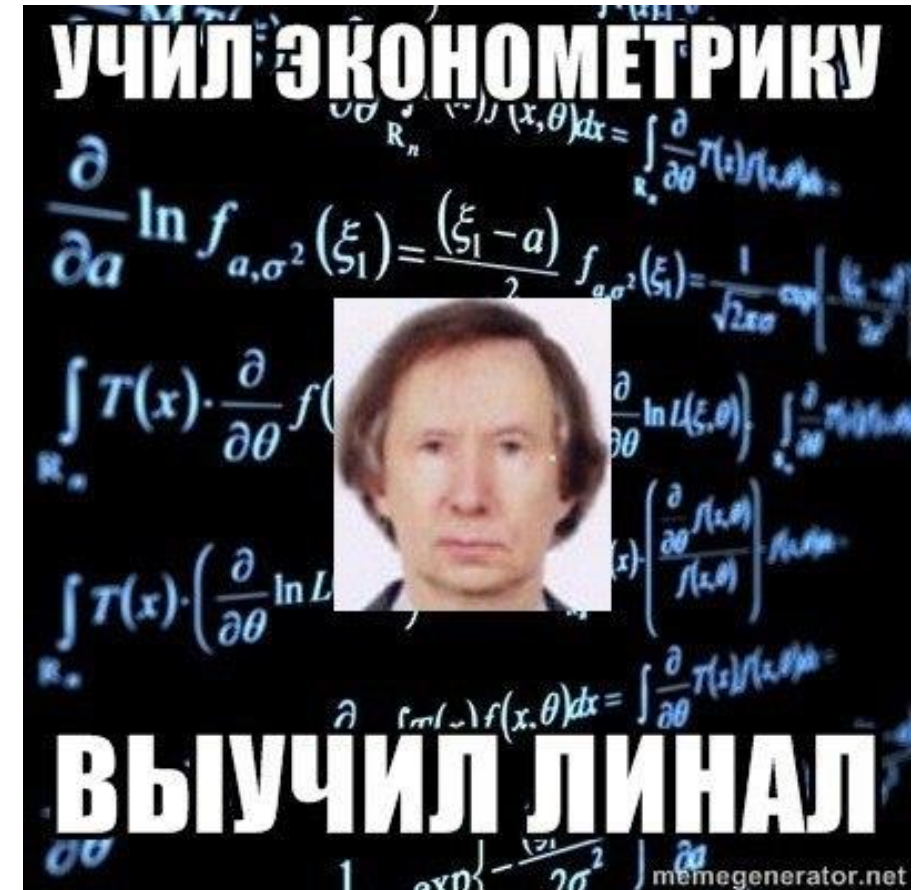
Царёв
Алексей

Формат общения

- Можно «Лёша» и нужно на «ты»;
- Вопросы можно во время лекции;
- Всё должно быть понятно. Если что-то непонятно, то лучше спросить;
- Глупых вопросов не бывает. Единственный человек, который говорит глупости всякие в этой аудитории – это я;
- Обязательно будет 15-минутный перерыв с пиццей и неофициальным общением;
- tsaryov.alexey.v@gmail.com – если есть вопросы, можно смело писать мне.

Содержание

- Основные понятия регрессионного анализа временных рядов;
- Метод скользящего среднего;
- Фильтр Ходрика-Прескотта;
- Фильтр Бакстер-Кинга;
- ES фильтр;
- ETS модели;
- Тета-метод;
- Эмпирика и реальные данные;
- Интерактив.



ОСНОВНЫЕ ПОНЯТИЯ

Временной ряд (Time Series) — последовательность случайных величин с восстановленной вероятностной мерой.

Макроэкономические показатели: ВВП, ВНП, ВРП, инфляция, безработица, номинальный обменный курс и т. д.

Финансовые данные: различные цены на финансовые инструменты, объёмы продаж и т. д.

Частота данных — это количество наблюдаемых значений временного ряда за фиксированный промежуток времени (годовые, квартальные, месячные, дневные данные).

Основные понятия

В динамике временного ряда могут наблюдаться систематические эффекты, а именно:

- некоторая основная тенденция, т. е. тренд (trend);
- циклические колебания (cyclicality);
- сезонная изменчивость (seasonality).

После учета этих систематических эффектов остается некоторая флуктуация, которая носит нерегулярный характер (irregularity).

В связи с этим, уровень ряда часто представляется:

- Либо в виде **суммы** тренда, циклической, сезонной и нерегулярной компонент – аддитивная модель, т. е.:

$$Y_t = T_t + C_t + S_t + I_t.$$

Основные понятия

- Либо в виде **произведения** этих компонент – мультипликативная модель:

$$Y_t = T_t * C_t * S_t * I_t.$$

Необходимо выделить и посчитать каждую из этих компонент.

Полученные оценки компонент можно использовать так и для построения новых временных рядов, как для прогнозирования.

Процедуры сглаживания «убирают» нерегулярную составляющую временного ряда так, что систематические эффекты становятся более явны.

Метод скользящего среднего

Пусть дан временной ряд y_t , $t = 1, \dots, T$ и некоторое число $K = 2m + 1$. Тогда сглаженный ряд $sma_i, i = m + 1, \dots, T - m$ образуется путём значения y_i в исходном ряде на значение:

$$sma_i = \frac{y_{i-m} + y_{i-m+1} + \dots + y_{i+m}}{K}.$$

Тогда ряд $c_i = y_i - sma_i$, $i = m + 1, \dots, T - m$ – это разность между фактическим значением ряда и сглаженным. Зачастую динамика данной компоненты не имеет систематического тренда.

Метод скользящего среднего

Проблема: а если K чётное? В макро рядах проявляются сезонные колебания (месячные, квартальные или полугодовые). Для таких рядов значения $K = 2, 4$ и 12 соответственно.

Решение: усредняем сначала сам ряд, а потом ещё раз его усредняем используя малое четное значение периода усреднения K , например, $K = 2$.

Полученный ряд sma_i^* называют **центрированным скользящим средним**, где:

$$sma_i^* = \frac{sma_{i-0,5} + sma_{i+0,5}}{2}.$$

Метод скользящего среднего

Пример $K = 4$. Если, $K = 4$, то какому периоду времени следует сопоставить усредненное значение:

$$\frac{(y_{i-1} + y_i + y_{i+1} + y_{i+2})}{4}$$

По аналогии со случаем нечетного K , это должен быть период $i + 0.5$; это порождало бы значения: $sma_{2.5}, sma_{3.5}, \dots, sma_{T-1.5}$.

В этом случае получаем в итоге последовательность значений: $sma_3^*, sma_4^*, \dots, sma_{T-2}^*$, где:

$$sma_i^* = \frac{sma_{i-0.5} + sma_{i+0.5}}{2}.$$

Заметим, что при этом:

$$sma_i^* = \frac{1}{4} (0.5y_{i-2} + (y_{i-1} + y_i + y_{i+1}) + 0.5y_{i+2}).$$

Выделение линейного тренда

Пусть имеется временной ряд y_t . Тогда выделить из него линейный тренд можно, построив следующую модель парной линейной регрессии:

$$y_t = \beta_0 + \beta_1 * t + \varepsilon_t.$$

Рассмотрим, например, задачу выделения долговременного тренда для определения разрыва ВВП.

Проблема: линейный тренд долгосрочный и не может меняться со временем.

Решение: нужно больше методов сглаживания!

Фильтр Ходрика-Прескотта

Предполагается, что ряд y_t представляется в виде:

$$y_t = y_t^{trend} + y_t^{cycle},$$

где y_t^{trend} – долговременный тренд, y_t^{cycle} – циклическая составляющая.

Ряд s_t подбираются таким образом, чтобы выполнялось условие:

$$\sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} ((s_{t+1} - s_t) - (s_t - s_{t-1}))^2 \rightarrow \min.$$

Полученный ряд s_t - оценка трендовой компоненты, $c_t = y_t - s_t$ - оценка циклической компоненты.

Что такое параметр λ ?

Как выбирать параметр λ ?

Ходрик и Прескотт для подбора λ использовали «правило степени 2»:

$$\lambda = \left(\frac{f}{4}\right)^2 * 1600 = f^2 * 100,$$

где f – частота ряда за год. Если f равен 1, 4 и 12, то λ равен 100, 1600 и 14400 соответственно. Но это не единственный способ:

Авторы статьи и год	Значение параметра λ	Частота данных
Ravn и Uhlig (2002)	6.5	Годовые
Kaiser и Maravall (1999)	8	Годовые
Pedersen (2001)	1000 и 3-5	Квартальные и годовые
Baxter и King (1995)	10	Годовые
Mohr (2001)	20	Годовые
Mohr (2001)	30	Годовые

Фильтр Ходрика-Прескотта

Ряд s_t может быть представлен как результат линейной фильтрации:

$$s_t = A(L)y_t,$$

где $A(L) = \sum_{s=-\infty}^{+\infty} a_s L^s$ – двусторонне скользящее среднее бесконечного порядка.

В 1993 году King и Rebelo показали, что оценка имеет s_t вид:

$$s_t = [1 + \lambda(1 - L)^2(1 - L^{-1})^2]^{-1}y_t = [1 + \lambda L^{-2}(1 - L)^4]^{-1}y_t.$$

Тогда оценка циклической компоненты имеет вид:

$$c_t = (1 - A(L))y_t = B(L)y_t = \frac{\lambda L^{-2}(1 - L)^4}{1 + \lambda L^{-2}(1 - L)^4}y_t.$$

Фильтр Ходрика-Прескотта

Пусть циклическая компонента содержит только колебания с частотами $\{\omega: 0 < \omega_0 \leq \omega \leq \pi\}$, что соответствует циклам $2\pi/\omega_0$ и менее. Тогда с учётом того, что в максимальной мере должны быть сохранены мощности включаемых в компоненту c_t колебаний, передаточная функция соответствующего линейного фильтра должна иметь вид:

$$|B(e^{-i\omega})| = \begin{cases} 0, & \text{если } |\omega| < \omega_0 \\ 1, & \text{если } \omega_0 \leq |\omega| < \pi \end{cases}$$

В свою очередь, сам фильтр $B(L)$ имеет вид:

$$B(L) = 1 - A(L) = 1 - \frac{\omega_0}{\pi} - \sum_{k=1}^{\infty} \frac{\sin(\omega_0 k)}{\pi k} (L^k + L^{-k}).$$

Фильтр Ходрика-Прескотта

Данный фильтр требует бесконечного числа наблюдений. Поэтому фильтры, построенные на любых конечных выборках, не могут отсекают заданные полосы частот, то есть не позволяют их выделить в полной мере.

Выбор Ходриком и Прескоттом для квартальных данных значения $\lambda = 1600$ был обусловлен желанием аппроксимировать идеальный фильтр верхних частот, дающий циклическую компоненту, которая включается в себя только случайные колебания с периодом менее 8 лет.

8 лет для квартальных данных – это 32 точки. Это соответствует частоте среза $\omega_0 = \frac{2\pi}{n} = \frac{\pi}{16} \approx 0,196$. 8 лет для годовых данных – это 8 точек. Это соответствует частоте среза $\omega_0 = \frac{2\pi}{n} = \frac{\pi}{4} \approx 0,785$.

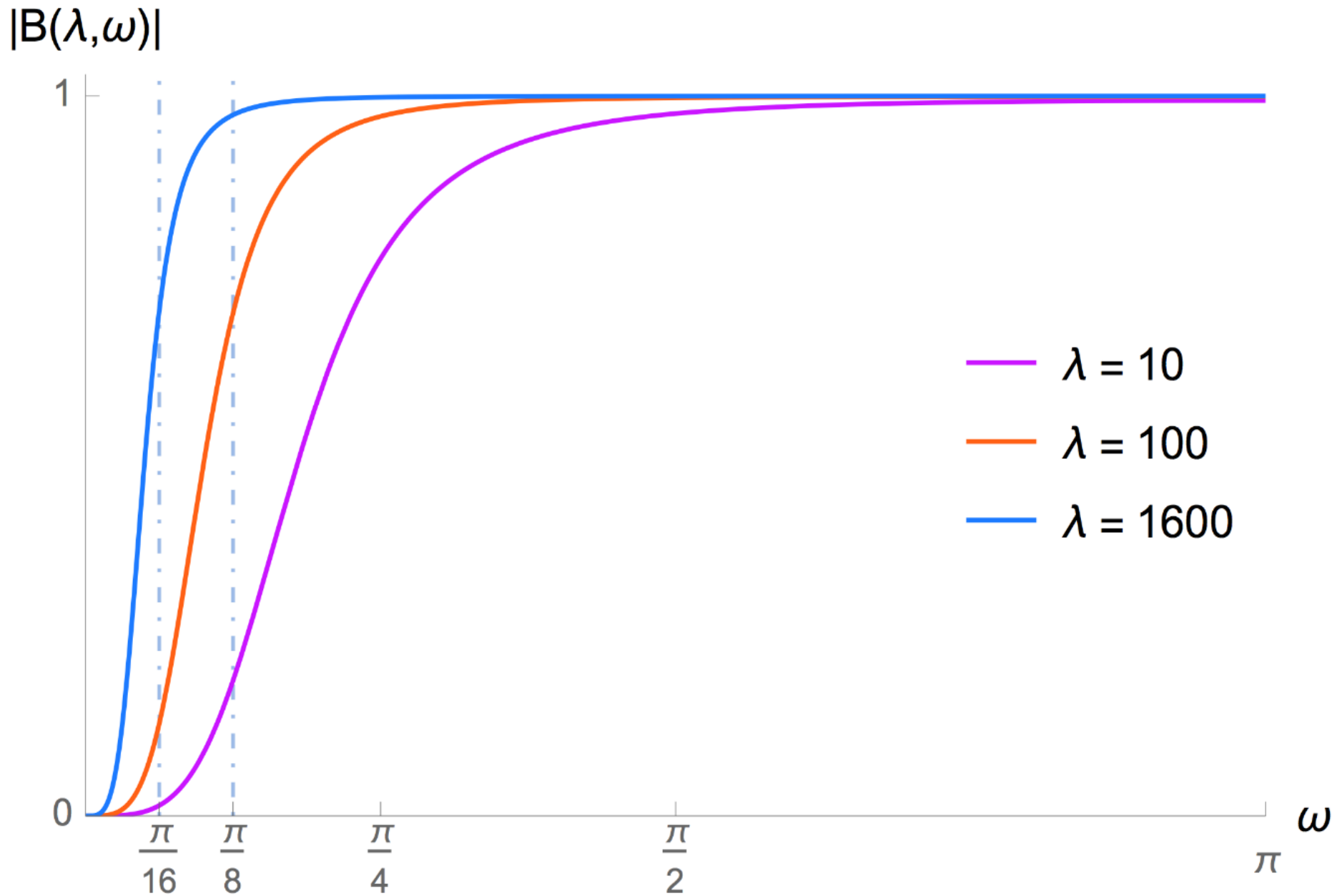
Фильтр Ходрика-Прескотта

Передающая функция фильтра Ходрика-Прескотта имеет вид:

$$|B(\lambda, \omega)| = \frac{4\lambda(1 - \cos(\omega))^2}{1 + 4\lambda(1 - \cos(\omega))^2}.$$

Заметим, что:

- $|B(\lambda, 0)| = 0$;
- $\lim_{\omega \rightarrow \infty} |B(\lambda, \omega)| = 1$;
- $\frac{\partial B}{\partial \omega} > 0$;
- $\exists \omega^*: \frac{\partial^2 B}{\partial \omega^2} = 0$



Передаточные функции фильтра Ходрика-Прескотта для параметров сглаживания 10, 100 и 1600

Фильтр Ходрика-Прескотта

Проблемы фильтра Ходрика-Прескотта на конечных выборках:

- Имеет место только сходимость последовательности передаточных функций к передаточной функции идеального фильтра некоторого (при правильной спецификации параметра λ).
- Индуцирование фазового сдвига. Однако, этим сдвигом можно пренебречь, за исключением первых и последних наблюдений в выборке. У этой проблемы есть 2 решения:
 1. Baxter и King в 1995 году предложили отбрасывать на обоих концах выборки получаемые значения c_t для 3 лет;
 2. Прогнозировании ряда за пределы имеющейся выборки и применении НР фильтра к расширенной выборке. В данном случае всё зависит от оптимальности прогноза.

Фильтр Ходрика-Прескотта

В 1983 году Whittle показал, что НР фильтр является оптимальным в смысле минимизации среднего квадрата ошибки, если:

$$\begin{cases} \Delta^2 y_t^{trend} = C(L)\zeta_t \\ y_t^{cycle} = C(L)\varepsilon_t \end{cases},$$

где $C(L) = \sum_{j=0}^{\infty} c_j L^j$, $\sum_{j=0}^{\infty} b_j^2 < \infty$ и $\zeta_t \sim NID(0, \sigma_\zeta^2)$, $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$. Тогда параметр сглаживания определяется как:

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}.$$

Для рядов $I(2)$:

- В 2000 году Reeves J.J. и др. показали, что оценка ряда s_t , получаемая в результате применения НР фильтра, является MLE.
- А в 2005 Schlicht получил выражение для логарифмической функции правдоподобия для параметра сглаживания λ .

Фильтр Ходрика-Прескотта

Hamilton (2016). Пусть $y_t \sim I(1)$, тогда ряд $(1 - L)y_t \sim I(0)$. В этом случае оценка циклической компоненты ряда будет равна:

$$c_t = \frac{\lambda L^{-2}(1 - L)^3}{1 + \lambda L^{-2}(1 - L)^4} (1 - L)y_t.$$

Чем плохи передифференцированные временные ряды:

1. Возникновение ложных циклических колебаний;
2. Долгосрочная дисперсия равна 0;
3. Не получится данный ряд вставить в модель;
4. Передифференцированные ряды не прогнозируемы.

Фильтр Бакстер-Кинга

БК фильтр (1995) – это попытка аппроксимировать с помощью центрированного скользящего среднего конечного порядка $M = 2K + 1$ идеальный полосовой фильтр, где флуктуации имеют период от 6 до 32 кварталов. То есть циклическая компонента представима в виде:

$$c_t = H_{BK}^K(L)y_t = \sum_{j=-K}^K a_j y_{t-j} = a_0 + \sum_{j=1}^K a_j (y_{t-j} + y_{t+j}).$$

Множество весов a_j получается путем обрезания весовых коэффициентов идеального фильтра при ограничении функции амплитудно-частотной характеристики:

$$H_0 = \Delta t \sum_{j=-\frac{N}{2}}^{\frac{N-1}{2}} a_j ,$$

где N – это количество наблюдений временного ряда y_t , а Δt – частота наблюдаемых значений временного ряда y_t . Передаточная функция БК фильтра имеет те же свойства, что и у НР фильтра.

Фильтр Бакстер-Кинга

В 2005 году Noullez и Iacobucci показали, что коэффициенты ВК фильтра являются решением следующей оптимизационной задачи:

$$\min_{a^{BK}} \int_{-(2\Delta t)^{-1}}^{(2\Delta t)^{-1}} \left| \left(\sum_{j=-K}^K a_j^{BK} - a_j^{ideal} \right) e^{-i2\pi n \omega \Delta t} \right|^2 d\omega.$$

Стоит отметить, что a_j^{ideal} вычисляются следующим образом:

$$a_j^{ideal} = \frac{\sin\left(\frac{2\pi j}{\underline{\omega}}\right) - \sin\left(\frac{2\pi j}{\overline{\omega}}\right)}{\pi j} \quad \forall j \geq 1,$$

где $\underline{\omega}$ и $\overline{\omega}$ – нижняя и верхняя частоты отсечения соответственно. Очевидно, что данное выражение неопределенно при $j = 0$, однако определён следующий предел:

$$\lim_{j \rightarrow 0} a_j^{ideal} = 2 \frac{\overline{\omega} - \underline{\omega}}{\underline{\omega} \overline{\omega}}.$$

Фильтр Бакстер-Кинга

Решение этой задачи показывает, что коэффициенты в ВК фильтре есть сумма :

$$a_j^{BK} = a_j^{ideal} + \frac{H(\omega) - \Delta t \sum_{j=-K}^K a_j^{ideal}}{M \Delta t} = a_j^{ideal} + \frac{H(\omega)}{M \Delta t} - \frac{\sum_{j=-K}^K a_j^{ideal}}{M}.$$

Стоит отметить, что:

- Коэффициенты ВК фильтра зависят от M и не зависят от N . Это означает, что качество построенного фильтра не изменяется по мере увеличения числа наблюдений.
- Существуют рекомендованные значения параметров $K, \underline{\omega}, \overline{\omega}$ в зависимости от частоты исходного ряда.

Таблица 1 – Рекомендованные значения параметров K , ω , $\bar{\omega}$ в зависимости от частоты исходного ряда

Частота ряда y_t	K	$\underline{\omega}$	$\bar{\omega}$
Годовая	3	2	8
Полугодовая	6	3	16
Квартальная	12	6	32
Месячная	36	8	96
Недельная	156	78	416

Фильтр Бакстер-Кинга

Свойства ВК фильтра:

- Благодаря симметричности фильтру несвойственен фазовый сдвиг в середине временного ряда, однако он возникает на концах выборки;
- Лаговый оператор $a(L)$ может быть приведён к следующему виду:

$$a(L) = -(1 - L)(1 - L^{-1})\psi_K(L),$$

где $\psi_K(L)$ – симметричное скользящее среднее с $K - 1$ запаздывающими и опережающими значениями.

- ВК фильтр способен приводить к стационарности ряды, содержащие вплоть до двух единичных корней.
- ВК фильтр не используется на концах ряда. Применение фильтра к временному ряду приводит к потере K наблюдений как в начале, так и в конце ряда. Низкие K – плохо, высокие – тоже плохо. Рекомендации: авторы предлагают принимать $K \geq 12$, не задумываясь о количестве наблюдений, частоте выборки и извлекаемой полосе. Используя $K = 12$, соответственно, будет означать потерю 24 наблюдений из ряда.

ETS-модели

ETS-модели – модели экспоненциального сглаживания. $ETS(*,*,*)$, где:

- E – Error;
- T – Trend Component;
- S – Seasonal Component;
- N – None, A – Additive, M - Multiplicative

$ETS(A,N,N)$ – простое экспоненциальное сглаживание; $ETS(A,A,N)$ – метод Хольта; $ETS(A,A,A)$ – метод Хольта-Уинтерса с аддитивной сезонностью; $ETS(A,A,M)$ – метод Хольта-Уинтерса с мультипликативной сезонностью.

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	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M_d	$y_t = \ell_{t-1}b_{t-1}^\phi + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

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

Простое экспоненциальное сглаживание

SES используется для сглаживания (прогнозирования) рядов, которые не имеют очевидного тренда или особой модели поведения. Сглаживание осуществляется образом:

$$s_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \dots + \alpha(1 - \alpha)^{t-1}y_1 = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i y_{t-i},$$

где $0 < \alpha \leq 1$ – параметр сглаживания. Чем ближе α к 0, тем большее значения придаётся более ранним данным. Чем ближе значение α к 1, тем больший вес имеют новые данные.

Параметр α можно подобрать по имеющимся данным, выбрав значение, минимизирующее сумму квадратов ошибок рекурсивных прогнозов.

Простое экспоненциальное сглаживание

Прогноз на период $t + 1$ будет иметь вид:

$$s_{t+1} = \alpha y_t + (1 - \alpha)s_t.$$

Отметим, что у SES есть ряд существенных недостатков:

- 1) $F_{t+h} = F_{t+1} = s_{t+1}, \forall h \in N$;
- 2) SES не учитывает возможность какого-либо функционального тренда (линейного, параболического, логарифмического и т.д.);
- 3) SES не учитывает возможность наличия в данных аддитивной или мультипликативной сезонности.

В дальнейших работах эти недостатки были учтены.

ES фильтр

Двустороннее скользящее среднее бесконечного порядка с экспоненциально убывающими коэффициентами (ES фильтр) является решением следующей задачи минимизации:

$$\sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=1}^{T-1} (\Delta s_{t+1})^2 \rightarrow \min.$$

Решением данной задачи является следующая фильтрация:

$$s_t = \left[\frac{\theta/\lambda}{1 - \theta^2} \right] \left[\sum_{j=0}^{\infty} \theta^j L^j + \sum_{j=0}^{\infty} \theta^j L^{-j} \right] y_t,$$

где $\theta = \frac{1+2\lambda-\sqrt{1+4\lambda}}{2\lambda} = \frac{((1+2\lambda)-\sqrt{(1+2\lambda)^2-4\lambda^2})}{2\lambda}$. Из данного равенства следует то, что $\theta < 1 \forall \lambda > 0$.

Таблица 2 – Значения параметра θ

Параметр сглаживания λ	Коэффициент θ
7.5	0.6955
15	0.7730
30	0.8333
60	0.8790
120	0.9128
240	0.9375
480	0.9554
960	0.9682

ES фильтр

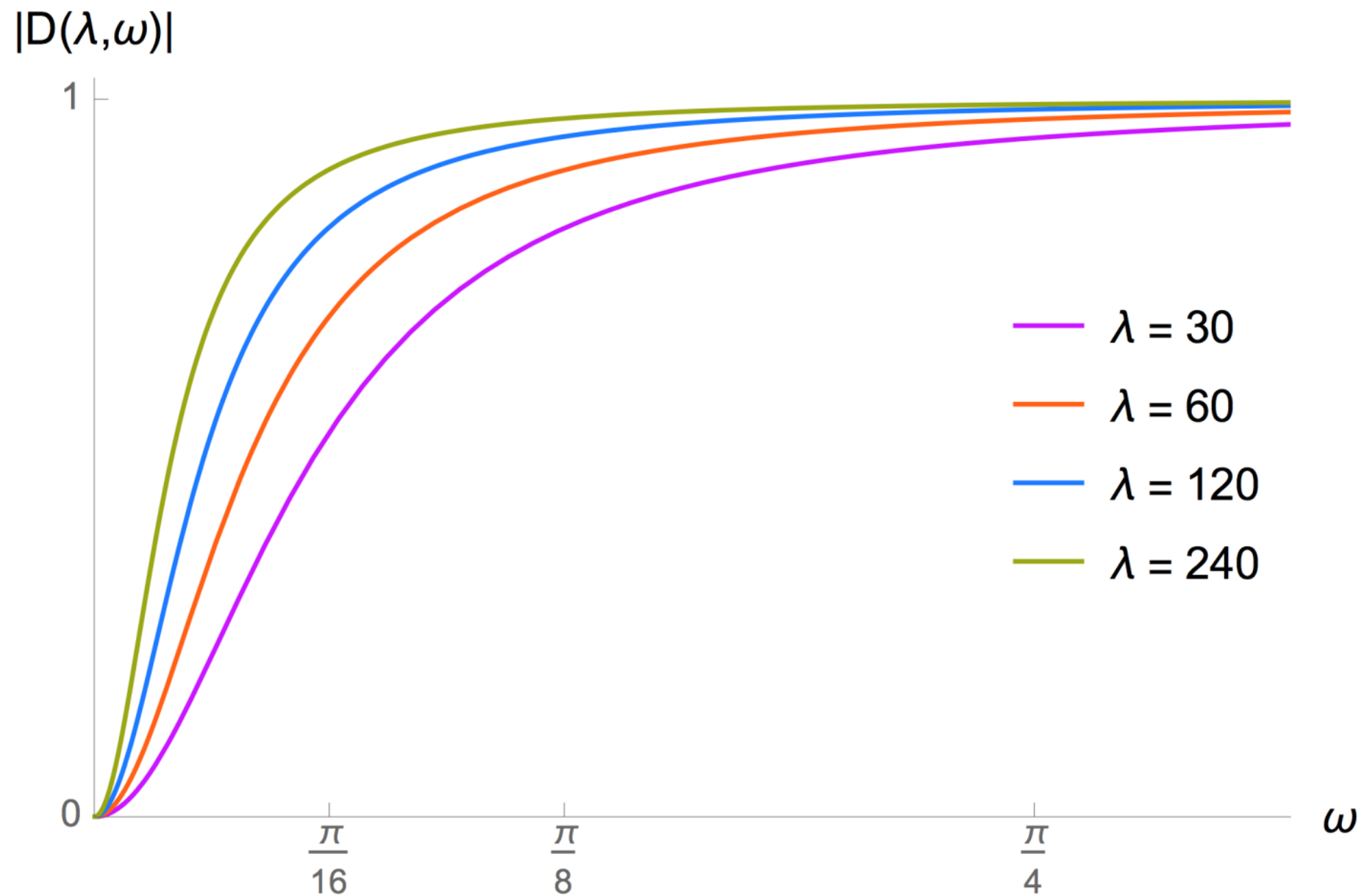
Оценка циклической компоненты c_t для ES фильтра имеет при этом вид:

$$c_t = \frac{(1 - L)(1 - L^{-1})}{1/\lambda + (1 - L)(1 - L^{-1})} y_t = \frac{\lambda L^{-1}(1 - L)^2}{1 + \lambda L^{-1}(1 - L)^2} y_t.$$

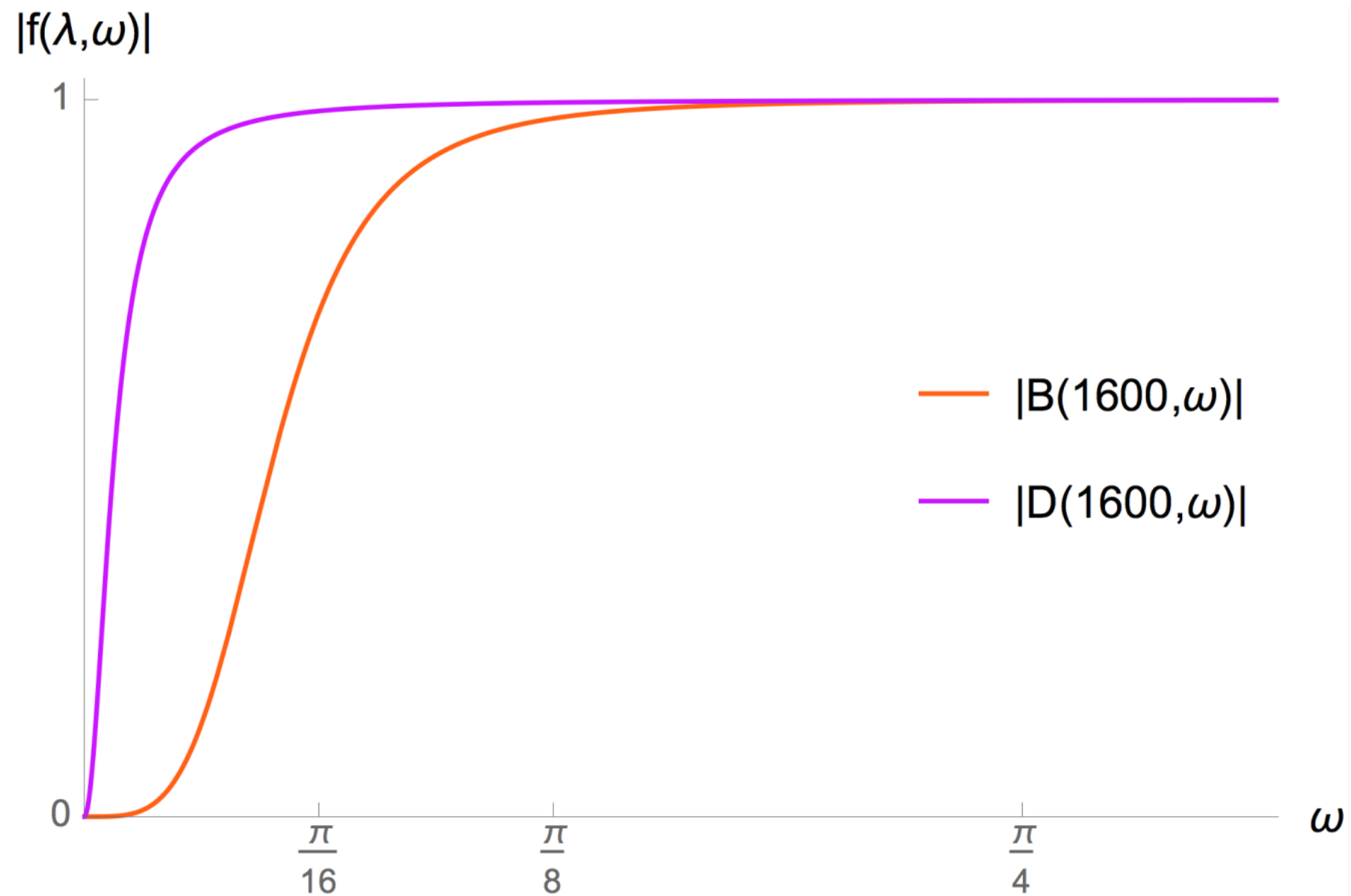
Передаточная функция ES фильтра имеет вид:

$$|D(\lambda, \omega)| = \frac{2\lambda(1 - \cos(\omega))}{1 + 2\lambda(1 - \cos(\omega))}.$$

Заметим, что у передаточной функции ES фильтра выполнены все те же самые свойства, что и у HP фильтра.



Передаточные функции ES фильтра для параметров сглаживания 30, 60, 120 и 240



Передаточные функций ES фильтра и HP фильтра для параметра сглаживания 1600

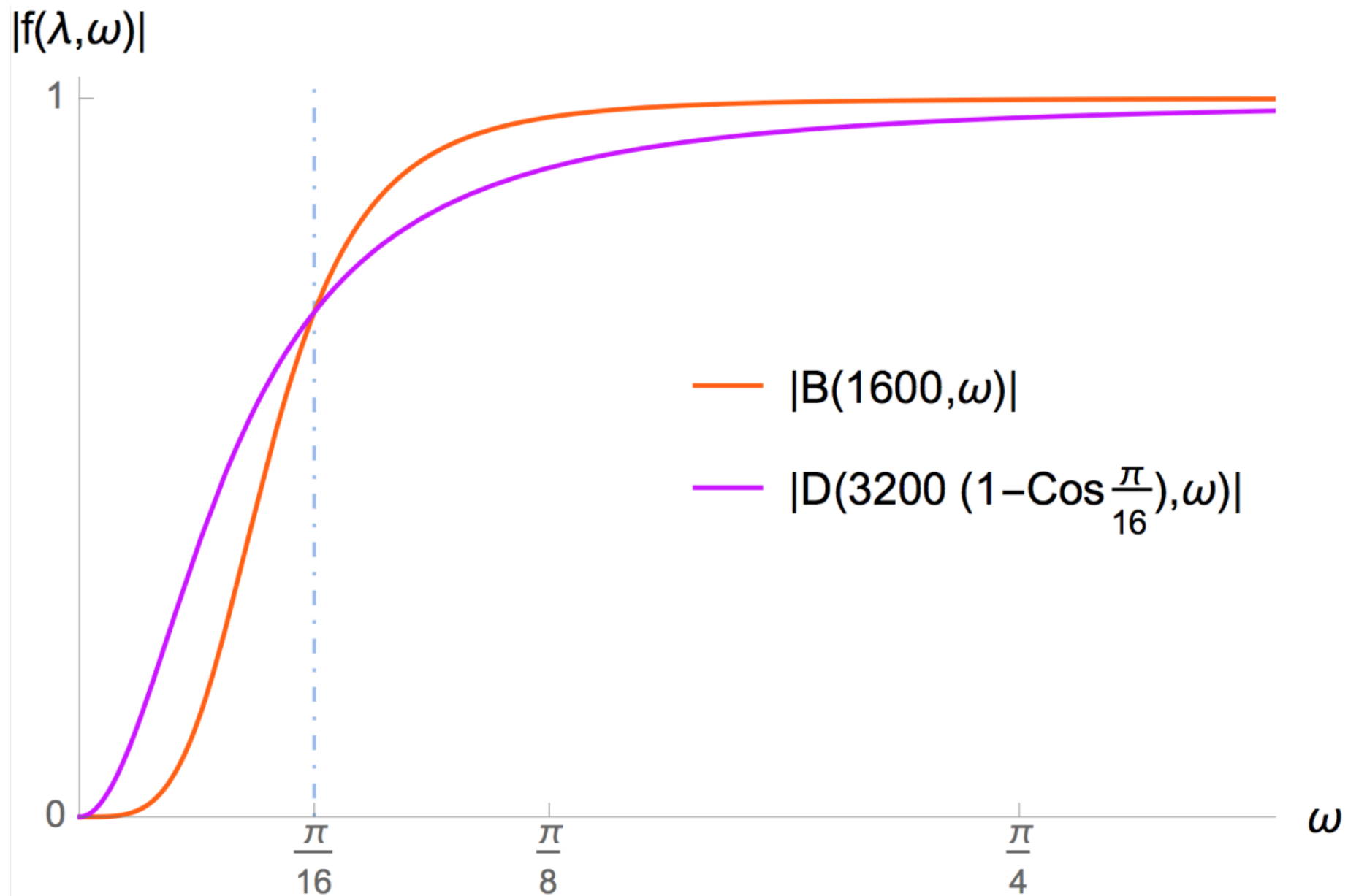
ES фильтр

Кривая передаточной функции для ES фильтра является более крутой, чем для фильтра Ходрика-Прескотта означает, что для обеспечения оптимального перераспределения частот можно взять более низкий параметр сглаживания. Оно определяется из равенства:

$$|B(\lambda(\omega^*), \omega^*)| = |D(\lambda, \omega^*)|$$

Например, для частоты $\omega^* = \frac{\pi}{16}$, $\lambda = 3200 \left(1 - \cos\left(\frac{\pi}{16}\right)\right) \approx 61.48$.

Из построенного далее графика можно сделать вывод о том, что ES фильтр лучше подходит для перераспределения более низких частот в сглаженном ряде.



Передаточные функций ES фильтра и НР фильтра для параметров сглаживания
61.48 и 1600 соответственно

ES фильтр

ES фильтр является оптимальным в смысле минимизации среднего квадрата ошибки, если:

$$\begin{cases} \Delta y_t^{trend} = C(L)\zeta_t \\ y_t^{cycle} = C(L)\varepsilon_t \end{cases},$$

где $C(L) = \sum_{j=0}^{\infty} c_j L^j$, $\sum_{j=0}^{\infty} c_j^2 < \infty$ и $\zeta_t \sim NID(0, \sigma_\zeta^2)$, $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$.

Тогда параметр сглаживания определяется как:

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}.$$

Метод Брауна

Метод Брауна стоит применять для рядов с выраженным линейным трендом, наличие которого учитывается при прогнозировании значений ряда на несколько периодов вперед. Метод Брауна – это двойное экспоненциальное сглаживание. То есть:

$$\begin{aligned}s_{t+1} &= \alpha y_t + (1 - \alpha)s_t, \\ s_{t+1}^* &= \alpha s_t + (1 - \alpha)s_t^*.\end{aligned}$$

Тогда:

$$F_{t+h} = \left(2 + \frac{\alpha h}{1 - \alpha}\right)s_t - \left(1 + \frac{\alpha h}{1 - \alpha}\right)s_t^* = (2s_t - s_t^*) + \frac{\alpha}{1 - \alpha}(s_t - s_t^*)h = a + bh,$$

где $a = 2s_t - s_t^*$, $b = \frac{\alpha}{1 - \alpha}(s_t - s_t^*)$ - параметры локального прямолинейного тренда. Прогнозные значения лежат на прямой, определяемой этими параметрами.

Метод Хольта

Хольт расширил SES, разработав модель, которая позволяет делать точный прогноз для данных с линейным трендом. Модель состоит из основного уравнения прогноза и двух сглаживающих уравнений, которые отвечают за уровень и тренд:

$$\begin{aligned}F_{t+h} &= l_t + hb_t, \\l_t &= \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}), \\b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}.\end{aligned}$$

Первое уравнение является основным в данной модели. Оно показывает сам прогноз значения временного ряда на период $t + h$, которое можно сделать на основе данных, имеющихся в момент времени t ; l_t отвечает за уровень, а b_t — за тренд.

Если в данных не наблюдается линейного тренда или имеет место сезонность, то эффективность использования этой модели для прогнозирования значительно снижается.

Метод Хольта-Уинтерса (аддитивная)

Аддитивная сезонность имеет место, когда сезонные отклонения данных примерно постоянны относительно изменяющегося среднего значения на протяжении всего наблюдения. В этом случае сезонная компонента выражается в абсолютных значениях и прибавляется к основному уравнению.

$$\begin{aligned}F_{t+h} &= l_t + hb_t + S_{t+h-n} \\l_t &= \alpha(y_t - S_{t-n}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\S_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)S_{t-n}\end{aligned}$$

Новое уравнение в модели (последнее), отвечающее за сезонность, является средневзвешенным между текущим индексом сезонности (в момент времени t) и индексом сезонности того же периода в прошлом году. Отсюда появляется значение $t - n$.

Метод Хольта-Уинтерса (мультипликативная)

Мультипликативная сезонность означает, что сезонное отклонение данных изменяется пропорционально изменению самих данных. И в этом случае сезонная компонента выражается в процентах (или в долях) и прогноз получается при помощи умножения основного уравнения на неё.

$$F_{t+h} = (l_t + hb_t)S_{t+h-n}$$
$$l_t = \alpha \frac{y_t}{S_{t-n}} + (1 - \alpha)(l_{t-1} + b_{t-1})$$
$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$
$$S_t = \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1 - \gamma)S_{t-n}$$

Тета-метод

Тета-метод (2000) заключается в том, что для прогнозирования исходного ряда Y_t рассчитывается такой ряд $s_{t,\theta}$, что для него выполняется следующее равенство:

$$\Delta^2 s_{t,\theta} = \theta * \Delta^2 y_t,$$

где $\Delta^2 y_t = y_{t+2} - 2y_{t+1} + y_t$. Полученный ряд $s_{t,\theta}$ называют тета-линией:

$$s_{t,\theta} = a_\theta + b_\theta(t - 1) + \theta y_t,$$

где a_θ и b_θ - константы.

Для первых двух элементов нового ряда $s_{t,\theta}$ находятся путём минимизации суммы квадратов расстояний до исходных данных:

$$\sum_{i=1}^t (y_i - s_{i,\theta})^2 = \sum_{i=1}^t [(1 - \theta)y_i - a_\theta - b_\theta(i - 1)]^2 \rightarrow \min_{a_\theta, b_\theta}$$

Тета-метод

Для прогнозирования используют средневзвешенный прогноз для $s_{t,\theta}$ при различных значениях θ . Авторы данного метода используют две тета-линии ($\theta = 0$ и $\theta = 2$), и тогда прогноз на основании n наблюдений на h периодов вперед выглядит следующим образом:

$$F_n(h) = \frac{1}{2} \left(\hat{s}_{n,0}(h) + \hat{s}_{n,2}(h) \right),$$

где $\hat{s}_{n,0}(h)$ представляет собой линейную часть уравнения $s_{t,\theta} = a_\theta + b_\theta(t - 1) + \theta y_t$ и является линейным трендом, рассчитанным на основе МНК, а $\hat{s}_{n,2}(h)$ получается при простом экспоненциальном сглаживании ряда $s_{t,2}$.



*То чувство, когда эконометрика
пригодилась в жизни.*

Перерыв 15 минут!