Stationary ARMA modeling and forecasting

(Updated Spring 2021)

Chaoyi Chen
Institute of MNB, Corvinus University of Budapest

Empirical Financial Econometrics

Outlines

- Brief Introduction to ARMA Models Jump (Online Lecture)
- Stationarity Jump (Self-study)
- Invertibility Jump (Self-study)

Brief Introduction to ARMA Models

- White Noise Process
- Uncorrelatdeness and weak exogeneity
- Moving Average Processes
- Autoregressive Models
- Autoregressive Moving average Processes–ARMA(p,q)models

White Noise Process

- "The basis building block"
- A white noise process ε_t satisfies:

$$
\begin{aligned}
E[\varepsilon_t] &= 0 && \text{for all t (mean zero)} && (1) \\
E[\varepsilon_t^2] &= \sigma^2 && \text{for all t (constant variances)} && (2) \\
Cov(\varepsilon_t \varepsilon_{t-s}) &= 0 && \text{for all t, all } s \neq 0 && (3)
\end{aligned}
$$

Relation between white noise process,normality and i.i.d process

- Assume ε_t has mean zero and variance σ^2

$$\begin{aligned} i.i.d &\Rightarrow whitenoise \quad (4) \\ \text{white noise} &\nRightarrow i.i.d \quad (5) \\ \text{white noise} + \text{normality} &\Rightarrow i.i.d \quad (6) \end{aligned}$$

Uncorrelatdeness and weak exogeneity

- Weak exogeneity $E_t \varepsilon_{t+1} = 0$
- Uncorrelatedness $E[\varepsilon_t \varepsilon_{t+s}] = 0, s \neq 0$
- **Weak exogeneity implies Uncorrelatedness**

  Proof:

  Suppose: $E_t \varepsilon_{t+1} = 0$, all t (weak exogeneity)

  Then: Letting $s > 0$

$$
\begin{aligned}
E[\varepsilon_t \varepsilon_{t+s}] &= E[E_{t+s-1}[\varepsilon_t \varepsilon_{(t+s)}]] && (7) \\
&= E[\varepsilon_t \underbrace{E_{t+s-1}[\varepsilon_{t+s}]}_{0 \text{ by weak exogeneity}}] && (8) \\
&= E[\varepsilon_t 0] = 0 && (9)
\end{aligned}
$$

Uncorrelatdeness and weak exogeneity

- This means that the assumptions

$$E_t \varepsilon_{t+1} = 0 \text{ and } E[\varepsilon_t^2] = \sigma^2$$

- also satisfy the white noise assumptions above

Intuition on white noise process:

- White noise process has no memory
- The name 'white noise' is due to its putting equal weight on cycles of all frequencies, similar to white noise or white light.
- It is a very simple building block
- It is easy to simulate on Matlab or even by repeatedly rolling dice or flipping a coin.
- For example: you could simulate white noise Bernoulli type Process by repeated coin toss:

  Just let t the t^{th} toss of the coin and define

$$\varepsilon_t = \begin{cases} 1 & \text{toss t is heads} \\ -1 & \text{toss t is tails} \end{cases}$$

Moving Average Processes: Example: Average Winnings

- Let ε_t be white noise

$$x_t = \sum_{i=0}^{q} b_i \varepsilon_{t-1} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \ldots + b_q \varepsilon_{t-q}$$

This called a Moving Average Process of order q or $MA(q)$ for short.

- **Examples:** A) Average winnings over two fair bets

$$\varepsilon_t = \begin{cases} 1 & \text{Heads (win)} \\ -1 & \text{Tails (lose)} \end{cases}$$

ε_t is your one period earnings
Your average earnings over the last two periods (dice rolls) is given by

$$x_t = \frac{1}{2}\varepsilon_t + \frac{1}{2}\varepsilon_{t-1}, \qquad MA(1)$$

- This is literally an average that moves, explaining the name

Moving Average Processes: Example Long-Horizon Returns

- **Example B: Long-horizon (log) stock returns**
  Recall that

$$r_t(k) = \sum_{i=0}^{k-1} r_{t-i} = r_t + r_{t-1} + \ldots r_{t-(k-1)}$$

- Suppose that the one-period log return is white noise
- Then the long horizon return follows a moving average:

$$r_t(k) \sim MA(k-1)$$

Moving Average Processes: Example long-horizon returns

- If we observe a white noise monthly (log) return then:
  - The quarterly return (3 months) is MA(2)
  - The yearly return (12) is MA(11)
- If we observe white noise daily log return then the weekly (5 day) return is MA(4)
- If we observe white noise weekly log return, then the monthly (4 week) return is MA(3)

Example: Shock whose impact gradually fades

- **Example C: Shock whose impact gradually fades (learning & forgetting)**

$$
\begin{aligned}
x_t &= \text{Number of French words I know at time } t \text{ (stock)} \\
\mu + \varepsilon_t &= \text{Number of new French words I learn at time t (flow)}
\end{aligned}
$$

- Suppose that I forget 1/3 of the new words after I just learned after one month, another 1/3 after two months, and the remaining third after three months.
- Let ε_t be white noise. Then my stock of french vocabulary follows an MA(2) with intercept:

$$
\begin{aligned}
x_t &= (\mu + \varepsilon_t) + \frac{2}{3}(\mu + \varepsilon_{t-1}) + \frac{1}{3}(\mu + \varepsilon_{t-2}) \\
&= 2\mu + \varepsilon_t + \frac{2}{3}\varepsilon_{t-1} + \frac{1}{3}\varepsilon_{t-2}
\end{aligned}
$$

Autoregressive Models

- Let ε_t be white noise and let

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t$$

Then we call y_t an **autoregressive process of order p** or $AR(p)$

Autoregressive Models: Oversimplified Examples

- **Examples** (Over-simplified, but illustrative)
  A) Wealth Accumulation

$$W_t = \text{Wealth at time t} \tag{10}$$
$$S_t = \text{Savings at time t} \tag{11}$$
$$i = \text{constant interest rate(over-simplified)} \tag{12}$$

$$W_t = (1+i)W_{t-1} + S_t$$

- Suppose that every year your target savings is $\bar{S}$, but you miss this target by an random amount(ε):

$$S_t = \bar{S} + \varepsilon_t$$

$$W_t = \underbrace{\bar{S}}_{a_0} + \underbrace{(1+i)}_{a_1} W_{t-1} + \varepsilon_t \qquad AR(1)$$

- **B) Capital Accumulation**
  Suppose each year your firm has a target investment of $\bar{I}$ and you miss this target by ε_t, which is white noise

$$\bar{I} = \text{target investment} \qquad I_t = \bar{I} + \varepsilon_t \text{ actual investment}$$
$$K_t = \text{Capital Stock at time t} \qquad \delta = \text{Depreciation rate}$$

Then the capital stock follows an $AR(1)$

$$K_t = K_{t-1} - \delta K_{t-1} + I_t \tag{13}$$
$$K_t = (1-\delta)K_{t-1} + \bar{I} + \varepsilon_t \tag{14}$$
$$\underbrace{K_t}_{y_t} = \underbrace{I}_{a_0} + \underbrace{(1-\delta)}_{a_1} \underbrace{K_{t-1}}_{y_{t-1}} + \varepsilon_t \tag{15}$$

Autoregressive Moving average Processes–$ARMA(p,q)$ models

$$
\begin{aligned}
\varepsilon_t &= WN(0,\sigma^2) \qquad \text{WN means the white noise here} && (16) \\
y_t &= \alpha_0 + \underbrace{\sum_{i=1}^{p} \alpha_i y_{t-1}}_{AR(p)} + \underbrace{\sum_{i=0}^{q} b_i \varepsilon_{t-i}}_{MA(q)} \qquad ARMA(p,q) && (17)
\end{aligned}
$$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}$$

Expressing MA, AR, and ARMA models using summation and lag notation. In brief:

$$a(L)y_t = b(L)\varepsilon_t$$

where

$$a(L) = 1 - \sum_{i=1}^{p} a_i L^i = 1 - a_1 L - a_2 L^2 - \ldots - a_p L^p$$

and

$$b(L) = 1 + \sum_{i=1}^{q} b_i L^i = 1 + b_1 L + b_2 L^2 + \ldots b_q L^q$$

At home: Rewrite MA, AR, and ARMA models first using summation notation and then using lag notation

Stationarity

- Stationary and Stationary restrictions
- Strict Stationary
- Covariance Stationary
- When does covariance Stationary holds?
- Moving Average Models Example
- Stationary of AR(1) when$|\alpha_1| < 1$
- Non-stationary Example: Random walk

- **Stationary and Stationary restrictions**
- Define $Y \underset{d}{=} X$ to mean that Y and X share the same CDF.[1]
- **Note**:
- This does **NOT** mean that $x = y$
- It does mean that x and y are drawn from the same distribution.
- And it means that any probability statement you make about Y is true for X and vice versa.
- e.g: If $P(0 \leq Y \leq 1) = \frac{1}{2}$,then $P(0 \leq X \leq 1) = \frac{1}{2}$

[1]Recall that this is Short for Cumulative Distribution Function, which is the function $F(x) = P(X < x)$ where P stands for probability

Stationary and Stationary restrictions

- We can apply the same concept to a vector of random variables.
- Let $\underset{k*1}{y} = (y_1, y_2, ..., y_k)'$ and $\underset{k*1}{x} = (x_1, x_2, ..., x_k)'$
- Then by $y \underset{d}{=} x$, we mean that x and y share the same joint CDF

- **Intuitive Definition**: y_t is strictly stationary if its distribution does not change over time.
- **Formal Definition**: The sequence y_t is strictly station if for all t and s.

$$y_t \underset{d}{=} y_{t+s} \tag{18}$$

And for all $t_1, t_2, ..., t_k$ and s

$$(y_{t1}, y_{t2}, y_{tk}) \overset{d}{=} (y_{t1+s}, y_{t2+s}, ..., y_{tk+s}) \tag{19}$$

Strict Stationary

- **Discussion** equation18 tells us that distribution of y_t doesn't change over time.
- Suppose we let $t_1 = t$ and $t_2 = t + 1$ and plug this into equation19. Then we get:

$$(y_t, t_t + 1) \underset{d}{=} (y_{t+s}, y_{t+1+s})$$

Which tells us that the joint distribution of (y_t, t_{t+1}) also doesn't change over time.

Covariance Stationary

- **Intuitive definition:** y_t is covariance stationary if the first 2 moments of y_t(the expectation, the variance and the covariance) don't change over time.
- **Formal Definition:** The sequence y_t is covariance stationary if for a t,h and s.

$$E[x_t] = E[x_{t+s}] \tag{20}$$
$$var(x_t) = var(x_{t+s}) \tag{21}$$
$$cov(x_t, x_{t+h}) = cov(x_{t+s}, x_{t+s+h}) \tag{22}$$

- **Discussion:**
  equation 22 says that $cov(x_t, x_{t+h})$ depends only on h, the distance in time between the two variables, but not on t.

When does covariance Stationary holds?

- **How to check Covariance Stationary?**
  1) calculate $E[y_t]$
  2) calculate $var[y_t]$
  3) calculate $cov(y_t, y_{t+h})$

  If any of these are infinite vary with t then y_t is **not** stationary. If they are finite and do not vary with t then y_t is Covariance Stationary.

Moving Average Models

- (A) MA(1) is always stationary

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

- **Step 1:**

$$E[y_t] = E[\varepsilon_t + \beta_1 \varepsilon_{t-1}] = \underbrace{E[\varepsilon_t]}_{0} + \beta_1 \underbrace{E[\varepsilon_{t-1}]}_{0} = 0,$$

which is finite and dot dependent on t

- **Step 2:**

$$\begin{aligned}
var(y_t) &= var(\varepsilon_t + \beta_1 \varepsilon_{t-1}) && (23)\\
&= var(\varepsilon_t) + var(\beta_1 \varepsilon_{t-1}) + 2cov(\varepsilon_t, \varepsilon_{t-1}) && (24)\\
&= \sigma^2 + \beta_1^2 \sigma^2 && (25)\\
&= (1 + \beta_1^2)\sigma^2 \qquad \text{finite \& not depending on t} && (26)
\end{aligned}$$

- **Step 3:** Calculate $cov(y_t, y_{t+h})$,$h = 1, 2, 3, ...$
- When $h = 1$

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim WN(0, \sigma^2) \tag{27}$$

$$\begin{aligned}
cov(y_t, y_{t+1}) &= cov(\varepsilon_t + \beta_1 \varepsilon_{t-1}, \varepsilon_{t+1} + \beta_1 \varepsilon_t) && (28) \\
&= cov(\varepsilon_t, \beta_1 \varepsilon_t) && (29) \\
&= \beta_1 cov(\varepsilon_t, \varepsilon_t) && (30) \\
&= \beta_1 var(\varepsilon_t) && (31) \\
&= \beta_1 \sigma^2 \quad \text{Finite and does not depend on t} && (32)
\end{aligned}$$

Moving Average Models

- $h = 2$

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim WN(0, \sigma^2) \tag{33}$$

$$cov(y_t, y_{t+2}) = cov(\varepsilon_t + \beta_1 \varepsilon_{t-1}, \varepsilon_{t+2} + \beta_1 \varepsilon_{t+1}) \tag{34}$$

$$= 0 \tag{35}$$

- $cov(y_t, y_{t+h}) = 0$ for $h = 3, 4, 5$, same reason.
- **Try at home:** Argue that MA(2) is covariance stationary.
- **General Rule:** Finite order moving average processes.i.e,MA(k) for any finite k, are always stationary.

Stationary of AR(1) when $|\alpha_1| < 1$

- Stationary AR(1) model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t \tag{36}$$
$$|\alpha_1| < 1 \tag{37}$$
$$E_{t-1}\varepsilon_t = 0 \tag{38}$$
$$var(\varepsilon_t) = \sigma^2 \tag{39}$$

Stationary of AR(1) when $|\alpha_1| < 1$

- To observe that this model is stationary, first convert it to its $MA(\infty)$ representation

$$
\begin{aligned}
y_t &= \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t && (40)\\
y_t - \alpha_1 y_{t-1} &= \alpha_0 + \varepsilon_t && (41)\\
(1 - \alpha_1 L) y_t &= \alpha_0 + \varepsilon_t && (42)\\
y_t &= \frac{1}{1 - \alpha_1 L}(\alpha_0 + \varepsilon_t) && (43)\\
&= \sum_{j=0}^{\infty} (\alpha_1 L)^j (\alpha_0 + \varepsilon_t) && (44)\\
&= \sum_{j=0}^{\infty} \alpha_1^j \underbrace{(L^j \alpha_0)}_{\alpha_0} + \sum_{j=0}^{\infty} \alpha_1^j \underbrace{L^j \varepsilon_t}_{\varepsilon_{t-j}} && (45)
\end{aligned}
$$

Stationary of AR(1) when $|\alpha_1| < 1$

- $MA(\infty)$ representation of $AR(1)$

$$y_t = \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j + \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{t-j} \tag{46}$$

$$= \frac{\alpha_0}{1-\alpha_1} + \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{t-j} \tag{47}$$

- Now confirm covariance stationary

$$E[y_t] = E[\underbrace{\frac{\alpha_0}{1-\alpha_1}}_{c} + \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{t-j}] \tag{48}$$

$$= \frac{\alpha_0}{1-\alpha_1} + \sum_{j=0}^{\infty} \alpha_1^j \underbrace{E[\varepsilon_{t-j}]}_{0} = \frac{\alpha_0}{1-\alpha_1} \tag{49}$$

Stationary of AR(1) when $|\alpha_1| < 1$

- And for the variance:

$$
\begin{aligned}
var[y_t] &= var[\underbrace{\frac{\alpha_0}{1-\alpha_1}}_{\text{c}} + \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{t-j}] && (50)\\
&= var\left(\sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{t-j}\right) && (51)\\
&= \sum_{j=0}^{\infty} (\alpha_1^j)^2 \underbrace{var(\varepsilon_{t-j})}_{\sigma^2} \qquad (\varepsilon_t \text{ uncorrleated}) && (52)\\
&= \sigma^2 \sum_{j=0}^{\infty} (\alpha_1^2)^j && (53)\\
&= \frac{\sigma^2}{1-\alpha_1^2} \qquad \text{(finite and not depend on t)} && (54)
\end{aligned}
$$

- same result for $cov(y_t, y_{t+h})$,but we will skip this to save time.

Non-stationary Example: Random walk

- Random walk model($AR(1)$ with $\alpha_1 = 1$)

$$
\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t, \quad t = 1, 2, 3, \ldots && (55) \\
y_0 &= 0 \quad \text{[Initialization]} && (56) \\
E_{t-1}\varepsilon_t &= 0 && (57) \\
var(\varepsilon_t) &= \sigma^2 && (58)
\end{aligned}
$$

Non-stationary Example: Random walk

- We calculate the variances as:

$$
\begin{aligned}
var(y_0) &= 0 && (59)\\
var(y_1) &= var(y_0+\varepsilon_1) = var(\varepsilon_1) = \sigma^2 && (60)\\
var(y_2) &= var(y_1+\varepsilon_2) = var(\varepsilon_1+\varepsilon_2) = 2\sigma^2 && (61)\\
var(y_3) &= 3\sigma^2 && (62)\\
&\vdots && (63)\\
var(y_t) &= t\sigma^2 \qquad \text{(variance depends on t,non-stationary)} && (64)
\end{aligned}
$$

- Note that if we did not initialize then $var(y_t)$ would be infinite

Non-stationary Example: Random Walk (continued)

- **Note:** The first difference of a random walk is white noise, which is both stationary and unpredictable white noise. Since

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

it follow that

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

- **Example:** If log stock price $\ln P_t$ follows a random walk then the return (excluding dividends),

$$r_t = \Delta \ln P_t$$

is white noise and unpredictable.

Relationship between Covariance and Strict Stationary

- $$\begin{aligned} \text{Strict Stationary} &\Rightarrow \text{Covariance Stationary} \quad (65) \\ \text{Covariance Stationary} &\not\Rightarrow \text{Strict Stationary} \quad (66) \\ \text{Covariance Stationary} + \text{Normality} &\Rightarrow \text{Strict Stationary} \quad (67) \end{aligned}$$

- **Intuition**
  The distribution determines the moments [equation65]
  But the first 2 moments do not by themselves determine the distribution[equation66]
  Except, if the data is normally distributed[equation 67]

Invertibility

- Invertibility
- Invertibility:MA(1) process
- Over-differenced process as non-invertible example
- Non-invertibly

Invertibility

1. **Intuitive Definition:** A process is invertible if it can be approximated by a finite order auto-regressive process.
2. We can estimate the AR(p) model by OLS so can estimate an invertible process
3. The MA(1) model
$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1}$$
has an AR(∞) representation of the form:[2]
$$y_t = -\sum_{j=0}^{\infty} \beta_1^j y_{t-j} + \varepsilon_t, \qquad AR(\infty) \text{ representation} \tag{68}$$
4. When $\beta < 1$ we drop the distant lags to approximate this by an AR(p)

[2] We derive this a few slides later

Invertibility (continued)

$$y_t = -\sum_{j=0}^{\infty} \beta_1^j y_{t-j} + \varepsilon_t, \qquad AR(\infty) \text{ representation} \tag{69}$$

1. However $\beta = 1$ the distant lags are too important to drop and the process

$$y_t = \varepsilon_t - \varepsilon_{t-1}$$

   is said to be over-differened.

2. This is because it differences an already stationary process

3. Below we provide a more precise technical definition and the derivation of equation (68)

Invertibility

- **Invertibility** An ARMA process is **invertible** if it can be expressed as AR process that is either **finite order** or **convergent**
- **Finite AR process:** An $AR(p)$,where p is finite.
- **Convergent AR process:**

$$y_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j y_{t-j} + \varepsilon_t, \qquad \text{where } \sum_{j=1}^{\infty} |\alpha_j| < \infty$$

- **Intuition:** We can approximate it by an AR(p) for finite p

- **Invertible MA(1) process:**

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} \quad |\beta_1| < 1 \quad \text{(Invertibility condition)} \tag{70}$$

$$E_{t-1}\varepsilon_t = 0 \qquad var(\varepsilon_t) = \sigma^2 \tag{71}$$

Can express this as a convergent $AR(\infty)$

$$y_t = \varepsilon_t - \beta_1 L \varepsilon_t \tag{72}$$

$$= (1 - \beta_1 L)\varepsilon_t \tag{73}$$

$$\frac{1}{1 - \beta_1 L} y_t = \varepsilon_t \tag{74}$$

$$\sum_{j=0}^{\infty} \beta_1^j L^j y_t = \varepsilon_t \tag{75}$$

$$\sum_{j=0}^{\infty} \beta_1^j y_{t-j} = \varepsilon_t \tag{76}$$

Invertibility:MA(1) process

Pull out the $j = 0$ term in the sum from the top of the last page:

$$\sum_{j=0}^{\infty} \beta_1^j y_{t-j} = \varepsilon_t \quad (77)$$

$$\beta_1^0 y_{t-0} + \sum_{j=1}^{\infty} \beta_1^j y_{t-j} = \varepsilon_t \quad (78)$$

$$y_t + \sum_{j=1}^{\infty} \beta_1^j y_{t-j} = \varepsilon_t \quad (79)$$

$$(80)$$

Solving for y_t it is recognizable as an AR(∞):

$$y_t = -\sum_{j=0}^{\infty} \beta_1^j y_{t-j} + \varepsilon_t, \qquad AR(\infty) \text{ representation} \quad (81)$$

Invertibility:MA(1) process

The MA(∞) representation:

$$y_t = -\sum_{j=0}^{\infty} \beta_1^j y_{t-j} + \varepsilon_t, \tag{82}$$

is a convergent AR(∞) because

$$\sum_{j=0}^{\infty} |\beta_1|^j = \frac{1}{1-|\beta_1|} < \infty, \qquad \text{for } |\beta_1| < 1$$

That means that the MA(1) is invertible

Over-differenced process as non-invertible example

- When $\beta_1 = 1$ the process is said to be over-differenced since

$$y_t = \varepsilon_t - \varepsilon_{t-1} = \Delta\varepsilon_t \quad (83)$$

  takes a first difference of a series that is already stationary, and thus not requiring differencing.

- It is an example of a **non-invertible** process. To convince yourself simply Substitute $\beta_1 = 1$ into the sum in the convergence condition:

$$\sum_{j=1}^{\infty} |\beta_1|^j = \sum_{j=1}^{\infty} 1 = \infty$$

- Since this sum is infinite the $AR(\infty)$ is **not convergent**
- Therefore the $MA(1)$ process with $|\beta_1| = 1$ is **not invertible**.

Non-invertibly

- Non-invertible cause estimation problems
- Intuitively, we cannot estimate an $MA(1)$ directly by a regression
- What if we could approximate it by AR(p) where p is not too large?
- If its invertible, then we can do this by omitting the terms for large j in the AR(∞)

$$y_t = -\sum_{j=0}^{\infty} \beta_1^j y_{t-j} + \varepsilon_t, \qquad AR(\infty) \text{ representation} \tag{84}$$

since β_1^j gets small when j gets large if $\beta < 1$

- But this is surely **not** the case when $\beta = 1$