

# General topics in non-linear time series analysis

Erik Lindström

Centre for Mathematical Sciences  
Lund University

LU/LTH & DTU

# General philosophy of the course

- ▶ The purpose of this course is primarily to make you better at modelling, not statistical theory, although you will get a bit of both.
- ▶ Implementing models is a (very good) way of understanding; *I will be more than happy to implement short segments of code during the lectures.*

DATA



SORTED



ARRANGED



PRESENTED  
VISUALLY



EXPLAINED  
WITH A STORY



# Overview

## Introduction

General questions

Example 1

Example 2

Non-linear models

## Linear model with time varying mean

## Transfer Functions

Volterra series

## Parametric non-linear models

SETAR

TARSO

STAR

Other nonlinear models

Comparion

IGAR

## Stability

# General

Extending beyond the scalar, linear, discrete time Gaussian models.

- ▶ Linear/Non-linear?
- ▶ Univariate/Multivariate?
- ▶ Discrete time/Continuous time?
- ▶ Parametric/Non-parametric?
- ▶ Time-invariant/ Time varying?

# General

Extending beyond the scalar, linear, discrete time Gaussian models.

- ▶ Linear/Non-linear?
- ▶ Univariate/Multivariate?
- ▶ Discrete time/Continuous time?
- ▶ Parametric/Non-parametric?
- ▶ Time-invariant/ Time varying?

# General

Extending beyond the scalar, linear, discrete time Gaussian models.

- ▶ Linear/Non-linear?
- ▶ Univariate/Multivariate?
- ▶ Discrete time/Continuous time?
- ▶ Parametric/Non-parametric?
- ▶ Time-invariant/ Time varying?

# General

Extending beyond the scalar, linear, discrete time Gaussian models.

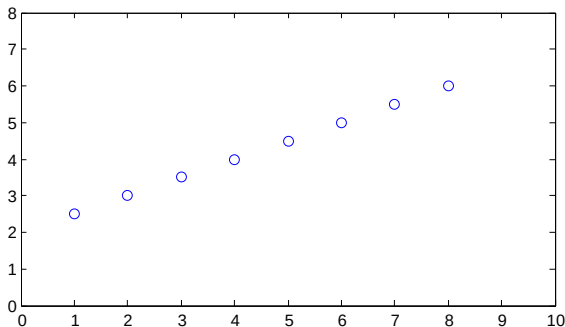
- ▶ Linear/Non-linear?
- ▶ Univariate/Multivariate?
- ▶ Discrete time/Continuous time?
- ▶ Parametric/Non-parametric?
- ▶ Time-invariant/ Time varying?

# General

Extending beyond the scalar, linear, discrete time Gaussian models.

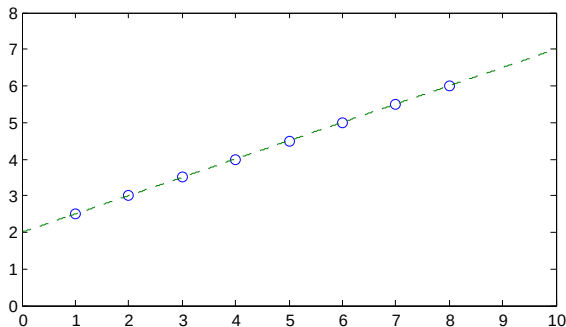
- ▶ Linear/Non-linear?
- ▶ Univariate/Multivariate?
- ▶ Discrete time/Continuous time?
- ▶ Parametric/Non-parametric?
- ▶ Time-invariant/ Time varying?

# New problems



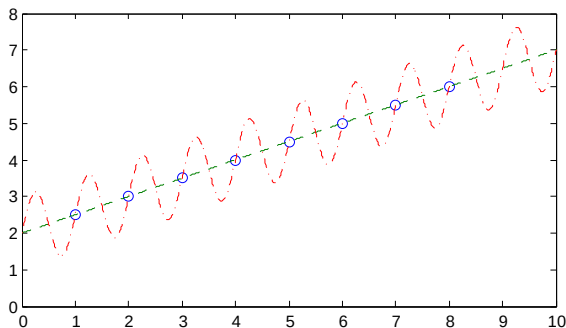
- Can you find a unique linear model?
- Is there a unique non-linear for the data?

# New problems



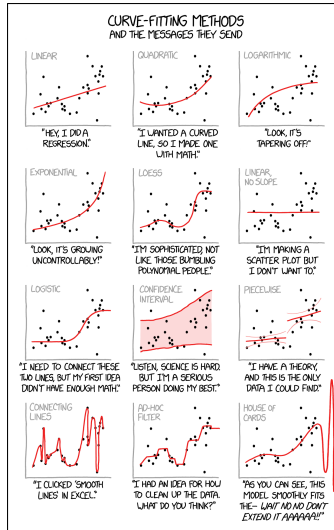
- Can you find a unique linear model? Yes
- Is there a unique non-linear for the data?

# New problems



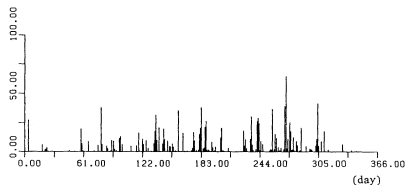
- Can you find a unique linear model? Yes
- Is there a unique non-linear for the data? No, several possible models!

# XKCD on regression models

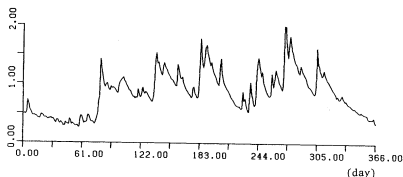


# River flow and rain fall in the Kanna river

This river flow is a non-linear process. Can you see that by using the definition of a linear model?



(5a)



(5b)

Figure 5. Kanna Rainfall and Riverflow Data, from January 1, 1956, to December 31, 1956.  
(5a) Kanna rainfall data; (5b) Kanna riverflow data.

# Non-linear features

- ▶ Limit cycles
- ▶ Jumps/Non-Gaussian sudden changes
- ▶ Skewed distributions
- ▶ Asymmetric responses
- ▶ Chaos?
- ▶ Bifurcations?

Adding a random component to the system tends to reduce the latter two.

# Non-linear features

- ▶ Limit cycles
- ▶ Jumps/Non-Gaussian sudden changes
- ▶ Skewed distributions
- ▶ Asymmetric responses
- ▶ Chaos?
- ▶ Bifurcations?

Adding a random component to the system tends to reduce the latter two.

# Tipping point

Recent, disturbing news

## nature communications

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature communications](#) > [articles](#) > article

Article | [Open Access](#) | [Published: 25 July 2023](#)

### Warning of a forthcoming collapse of the Atlantic meridional overturning circulation

[Peter Ditlevsen](#)  & [Susanne Ditlevsen](#) 

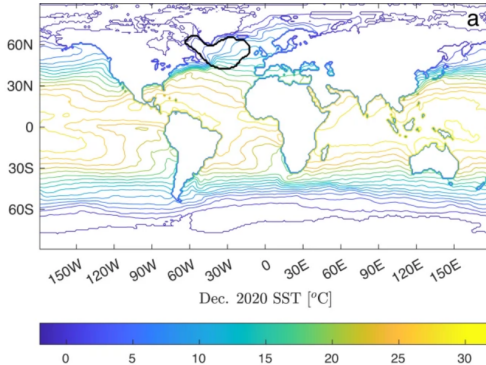
[Nature Communications](#) **14**, Article number: 4254 (2023) | [Cite this article](#)

**309k** Accesses | **1** Citations | **6050** Altmetric | [Metrics](#)

Predicts tipping point of the AMOC will occur with 95 % probability during 2025 – 2095 with a mode at 2057.

# Tipping point

**Fig. 1: The Atlantic meridional overturning circulation (AMOC) fingerprint, sea surface temperature (SST) and global mean (GM).**



Measurements of Sea Surface Temperature, SST, indicates that the AMOC is weakening.

# Review: Difference equations

Difference equation representation for ARX/ARMA structure:

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = u_t + b_1 u_{t-1} + \dots + b_q u_{t-q}$$

using the delay operator leads to the *Transfer Function*  $H(\cdot)$

$$y_t = H(z)u_t = \frac{1 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} u_t$$

with (the latter equation with a Z-transform interpretation of the operations). Note that transfer functions is more general than linear difference equations

# Review: Difference equations

Difference equation representation for ARX/ARMA structure:

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = u_t + b_1 u_{t-1} + \dots + b_q u_{t-q}$$

using the delay operator leads to the *Transfer Function*  $H(\cdot)$

$$y_t = H(z)u_t = \frac{1 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} u_t$$

with (the latter equation with a Z-transform interpretation of the operations). Note that transfer functions is more general than linear difference equations

## ARMA(p,q)-filter

Process:

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = x_t + c_1 x_{t-1} + \dots + c_q x_{t-q}$$

Transfer Function:

$$\begin{aligned} H(z) &= \frac{1 + c_1 z^{-1} + \dots + c_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} \\ &= \frac{z^{-q}(z^q + c_1 z^{q-1} + \dots + c_q)}{z^{-p}(z^p + a_1 z^{p-1} + \dots + a_p)} \end{aligned}$$

Stability: Poles  $|\pi_i| < 1, i = 1, \dots, p$

Invertability: Zeroes  $|\eta_i| < 1, i = 1, \dots, q$

# Frequency representation

The *Frequency function* is defined from the transfer function as

$$\mathcal{H}(f) = H(e^{i2\pi f}), f \in (-\pi, \pi]$$

giving a amplitude and phase shift of an input trigonometric signal, as e.g.

$$u_k = \cos(2\pi fk)$$

$$y_k = |\mathcal{H}(f)| \cos(2\pi fk + \arg(\mathcal{H}(f)))$$

$$|f| \leq 0.5$$

# What about non stationary linear models

- ▶ Variance depends on level - Box-Cox transformation
- ▶ Use differentiation if the data is non-stationary OR find a model that includes the non-stationary component as an integral (exogenous) part of the model
- ▶ Time varying mean - add deterministic function

$$\xi_t = \underbrace{y_t}_{\text{ARMAX}} + \underbrace{f_t}_{\text{Time-varying function}} \quad (1)$$

**Question:** Is it equivalent to first estimate  $f_t$ , subtract it, and then estimate parameters in  $y_t$  compared to estimating all parameters simultaneously?

- ▶ Non-Gaussian distribution (e.g. Heavy tails) - Change distribution in the model from Gaussian to say student- $t$ , Laplace, VG or NIG.

# What about non stationary linear models

- ▶ Variance depends on level - Box-Cox transformation
- ▶ Use differentiation if the data is non-stationary OR find a model that includes the non-stationary component as an integral (exogenous) part of the model
- ▶ Time varying mean - add deterministic function

$$\xi_t = \underbrace{y_t}_{\text{ARMAX}} + \underbrace{f_t}_{\text{Time-varying function}} \quad (1)$$

**Question:** Is it equivalent to first estimate  $f_t$ , subtract it, and then estimate parameters in  $y_t$  compared to estimating all parameters simultaneously?

- ▶ Non-Gaussian distribution (e.g. Heavy tails) - Change distribution in the model from Gaussian to say student- $t$ , Laplace, VG or NIG.

# What about non stationary linear models

- ▶ Variance depends on level - Box-Cox transformation
- ▶ Use differentiation if the data is non-stationary OR find a model that includes the non-stationary component as an integral (exogenous) part of the model
- ▶ Time varying mean - add deterministic function

$$\xi_t = \underbrace{y_t}_{\text{ARMAX}} + \underbrace{f_t}_{\text{Time-varying function}} \quad (1)$$

**Question:** Is it equivalent to first estimate  $f_t$ , subtract it, and then estimate parameters in  $y_t$  compared to estimating all parameters simultaneously?

- ▶ Non-Gaussian distribution (e.g. Heavy tails) - Change distribution in the model from Gaussian to say student- $t$ , Laplace, VG or NIG.

# What about non stationary linear models

- ▶ Variance depends on level - Box-Cox transformation
- ▶ Use differentiation if the data is non-stationary OR find a model that includes the non-stationary component as an integral (exogenous) part of the model
- ▶ Time varying mean - add deterministic function

$$\xi_t = \underbrace{y_t}_{ARMAX} + \underbrace{f_t}_{\text{Time-varying function}} \quad (1)$$

**Question:** Is it equivalent to first estimate  $f_t$ , subtract it, and then estimate parameters in  $y_t$  compared to estimating all parameters simultaneously?

- ▶ Non-Gaussian distribution (e.g. Heavy tails) - Change distribution in the model from Gaussian to say student- $t$ , Laplace, VG or NIG.

# What about non stationary linear models

- ▶ Variance depends on level - Box-Cox transformation
- ▶ Use differentiation if the data is non-stationary OR find a model that includes the non-stationary component as an integral (exogenous) part of the model
- ▶ Time varying mean - add deterministic function

$$\xi_t = \underbrace{y_t}_{\text{ARMAX}} + \underbrace{f_t}_{\text{Time-varying function}} \quad (1)$$

**Question:** Is it equivalent to first estimate  $f_t$ , subtract it, and then estimate parameters in  $y_t$  compared to estimating all parameters simultaneously?

- ▶ Non-Gaussian distribution (e.g. Heavy tails) - Change distribution in the model from Gaussian to say student- $t$ , Laplace, VG or NIG.

# Transfer functions for non-linear systems

- ▶ Linear models are often characterised by their transfer function.
- ▶ General model: For a given stationary time series  $\{X_t\}$  find a function  $h$  such that

$$h(\{X_t\}) = \epsilon_t \quad (2)$$

where  $\epsilon_t$  is a white noise sequence. Example. A linear model is defined as

$$\sum_{k=-\infty}^{\infty} h_{t,k} X_{t-k} = \epsilon_t. \quad (3)$$

- ▶ A model is globally *invertible* if it is possible to compute  $\{X_t\}$  from  $\{\epsilon_t\}$  and a given initial value.

# Transfer functions for non-linear systems

- ▶ Linear models are often characterised by their transfer function.
- ▶ General model: For a given stationary time series  $\{X_t\}$  find a function  $h$  such that

$$h(\{X_t\}) = \epsilon_t \quad (2)$$

where  $\epsilon_t$  is a white noise sequence. Example. A linear model is defined as

$$\sum_{k=-\infty}^{\infty} h_{t,k} X_{t-k} = \epsilon_t. \quad (3)$$

- ▶ A model is globally *invertible* if it is possible to compute  $\{X_t\}$  from  $\{\epsilon_t\}$  and a given initial value.

# Transfer functions for non-linear systems

- ▶ Linear models are often characterised by their transfer function.
- ▶ General model: For a given stationary time series  $\{X_t\}$  find a function  $h$  such that

$$h(\{X_t\}) = \epsilon_t \quad (2)$$

where  $\epsilon_t$  is a white noise sequence. Example. A linear model is defined as

$$\sum_{k=-\infty}^{\infty} h_{t,k} X_{t-k} = \epsilon_t. \quad (3)$$

- ▶ A model is globally *invertible* if it is possible to compute  $\{X_t\}$  from  $\{\epsilon_t\}$  and a given initial value.

# Volterra series

Start with *causal* model

$$h(X_t, X_{t-1}, \dots) = \epsilon_t \quad (4)$$

and assume that it is *causally invertible*

$$X_t = h'(\epsilon_t, \epsilon_{t-1}, \dots). \quad (5)$$

We then assume  $h'$  is sufficiently regular so that it can be expanded in a Taylor series.

# Volterra series

$$X_t = \mu + \sum_{k=0}^{\infty} g_k \epsilon_{t-k} + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g_{kl} \epsilon_{t-k} \epsilon_{t-l} \quad (6)$$

$$+ \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_{klm} \epsilon_{t-k} \epsilon_{t-l} \epsilon_{t-m} + R_t \quad (7)$$

where

$$\mu = h'(0, 0, \dots), \quad g_k = \frac{\partial h'}{\partial \epsilon_{t-k}}, \quad g_{kl} = \frac{\partial^2 h'}{\partial \epsilon_{t-k} \partial \epsilon_{t-l}} \quad (8)$$

# Volterra series

Let  $U_t$  be an arbitrary input signal, and define  $X_t$  as follows

$$X_t = \mu + \sum_{k=0}^{\infty} g_k U_{t-k} + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g_{kl} U_{t-k} U_{t-l} \quad (9)$$

$$+ \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_{klm} U_{t-k} U_{t-l} U_{t-m} \quad (10)$$

The transfer function for a linear system is given by the first two terms

$$X_t = \mu + \sum_{k=0}^{\infty} g_k U_{t-k} \quad (11)$$

The output, when the input is given by  $U_t = A_0 e^{i\omega_0 t}$  is a single harmonic with the *same* frequency, scaled by  $|H(\omega_0)|$  and phase shifted by  $\arg H(\omega_0)$ .

# Volterra series

This is not true for general non-linear systems. We have that

- ▶ For an input with frequency  $\omega_0$ , the output will also contain  $2\omega_0, 3\omega_0, \dots$
- ▶ For two input frequencies,  $\omega_0$  and  $\omega_1$ , the output will also contain frequencies  $\omega_0, \omega_1, \omega_0 + \omega_1$  and all harmonics of the frequencies.

# Volterra series

There is no such thing as a transfer function for non-linear systems. Instead, there is an infinity sequence of *generalized transfer functions*

$$H_1(\omega_1) = \sum_{k=0}^{\infty} g_k e^{-i\omega_1 k} \quad (12)$$

$$H_2(\omega_1, \omega_2) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g_{kl} e^{-i(\omega_1 k + \omega_2 l)} \quad (13)$$

$$H_3(\omega_1, \omega_2, \omega_3) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_{klm} e^{-i(\omega_1 k + \omega_2 l + \omega_3 m)} \quad (14)$$

# Different strategies for non-linear modelling

## Various strategies

- ▶ Using only external information from e.g. physics to describe the system. This type of white box models are not considered in the course.
- ▶ Black box models, such as SETAR, STAR, FAR, BL, Volterra series, random forests or (Deep) Neural Nets.
- ▶ Grey box models, combining sources of information about the system to be modelled (cf. Scientific Machine Learning - SciML). This is cover later in this course

# Different strategies for non-linear modelling

## Various strategies

- ▶ Using only external information from e.g. physics to describe the system. This type of white box models are not considered in the course.
- ▶ Black box models, such as SETAR, STAR, FAR, BL, Volterra series, random forests or (Deep) Neural Nets.
- ▶ Grey box models, combining sources of information about the system to be modelled (cf. Scientific Machine Learning - SciML). This is cover later in this course

# Different strategies for non-linear modelling

## Various strategies

- ▶ Using only external information from e.g. physics to describe the system. This type of white box models are not considered in the course.
- ▶ Black box models, such as SETAR, STAR, FAR, BL, Volterra series, random forests or (Deep) Neural Nets.
- ▶ Grey box models, combining sources of information about the system to be modelled (cf. Scientific Machine Learning - SciML). This is cover later in this course

# A word of caution from a grumpy old man...

From M3 forecasting competition. Forecasting performance across multiple forecasting horizons using a large subset of 1045 monthly time series

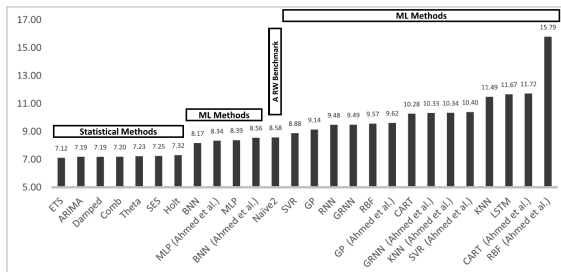


Fig 2. Forecasting performance (sMAPE) of the ML and statistical methods included in the study. The results are reported for one-step-ahead forecasts having applied the most appropriate preprocessing alternative.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.

# A word of caution from a grumpy old man...

From M3 forecasting competition. Forecasting performance across multiple forecasting horizons using a large subset of 1045 monthly time series

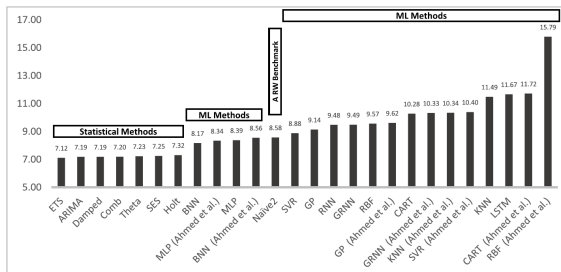


Fig 2. Forecasting performance (sMAPE) of the ML and statistical methods included in the study. The results are reported for one-step-ahead forecasts having applied the most appropriate preprocessing alternative.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.

# SETAR models

Define disjoint regions  $R_1, \dots, R_l$ , typically  $R_i = (r_{i-1}, r_i]$ . The values  $r_0, \dots, r_l$  are called thresholds.

The  $SETAR(l, d, k_1, \dots, k_l)$  model is defined as:

$$X_t = a_0^{(J_t)} + \sum_{i=1}^{k_{J_t}} a_i^{(J_t)} X_{t-i} + \epsilon_t^{(J_t)} \quad (15)$$

where

$$J_t = \begin{cases} 1 & \text{if } X_{t-d} \in R_1 \\ 2 & \text{if } X_{t-d} \in R_2 \\ \vdots & \vdots \\ l & \text{if } X_{t-d} \in R_l \end{cases} \quad (16)$$

Note: Similar to linear splines.

# Example SETAR model

Exchange rate, interest rate or *inflation corridors*.

## Historical background

- ▶ Inflation was (very) high in Sweden during the 70s and 80s.
- ▶ High inflation is often volatile. This worries investors and affects the economy negatively
- ▶ (Too) low inflation may lead to deflation, with negative incentives. This also affects the economy negatively
- ▶ Inflation target of  $2\% \pm 1\%$  set in 1993, effective from 1st January 1995.

# Example SETAR model

Exchange rate, interest rate or *inflation corridors*.

## Historical background

- ▶ Inflation was (very) high in Sweden during the 70s and 80s.
- ▶ High inflation is often volatile. This worries investors and affects the economy negatively
- ▶ (Too) low inflation may lead to deflation, with negative incentives. This also affects the economy negatively
- ▶ Inflation target of  $2\% \pm 1\%$  set in 1993, effective from 1st January 1995.

# Example SETAR model

Exchange rate, interest rate or *inflation corridors*.

## Historical background

- ▶ Inflation was (very) high in Sweden during the 70s and 80s.
- ▶ High inflation is often volatile. This worries investors and affects the economy negatively
- ▶ (Too) low inflation may lead to deflation, with negative incentives. This also affects the economy negatively
- ▶ Inflation target of  $2\% \pm 1\%$  set in 1993, effective from 1st January 1995.

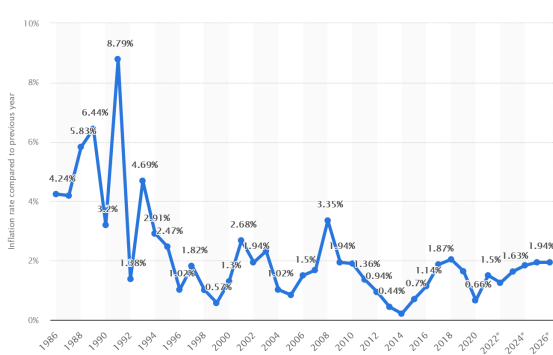
# Example SETAR model

Exchange rate, interest rate or *inflation corridors*.

## Historical background

- ▶ Inflation was (very) high in Sweden during the 70s and 80s.
- ▶ High inflation is often volatile. This worries investors and affects the economy negatively
- ▶ (Too) low inflation may lead to deflation, with negative incentives. This also affects the economy negatively
- ▶ Inflation target of  $2\% \pm 1\%$  set in 1993, effective from 1st January 1995.

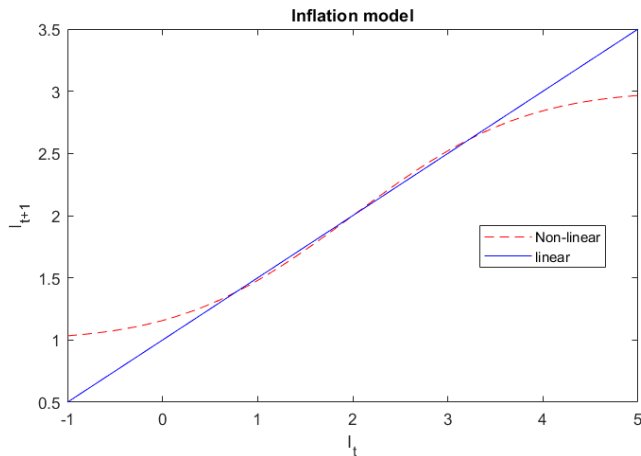
# Swedish inflation 1986-2021



Overly simplistic model.

- ▶  $I_{t+1} = a + bI_t + e_{t+1}$  or
- ▶  $I_{t+1} = f(I_t) + e_{t+1}$

Comparison of models for the inflation rate - is the policy enforced in practice?

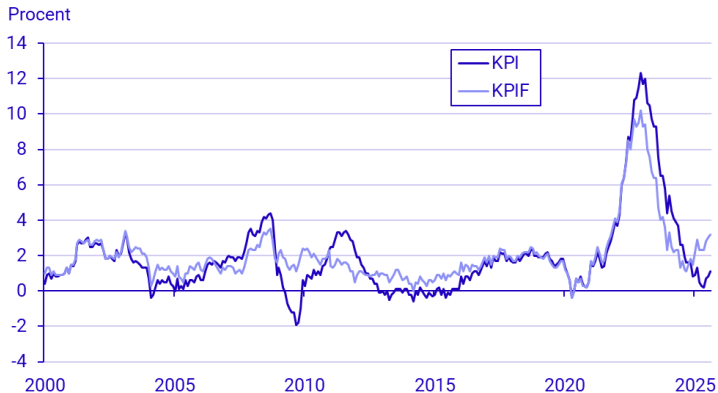


# Updated inflation data (-July 2023)



Did the central bank succeed in keeping inflation within the desired inflation corridor?

# Updated inflation data (-July 2026)



Did the central bank succeed in keeping inflation within the desired inflation corridor?

# TARSO models

Extends the SETAR model by allowing for **external signals**  $U_t$ , which is also the switching variable.

The  $TARSO(l, d, (k_1, k'_1) \dots, (k_l, k'_l))$  model is defined as:

$$X_t = a_0^{(J_t)} + \sum_{i=1}^{k_{J_t}} a_i^{(J_t)} X_{t-i} + \sum_{i=1}^{k'_{J_t}} b_i^{(J_t)} U_{t-i} + \epsilon_t^{(J_t)} \quad (17)$$

where

$$J_t = \begin{cases} 1 & \text{if } U_{t-d} \in R_1 \\ 2 & \text{if } U_{t-d} \in R_2 \\ \vdots & \vdots \\ l & \text{if } U_{t-d} \in R_l \end{cases} \quad (18)$$

# Öresund: open water or frozen?



# STAR models

Continuous alternative to SETAR (can estimate the thresholds)

The  $STAR(d, p)$  model is defined as:

$$X_t = a_0 + \sum_{j=1}^p a_j X_{t-j} + I(X_{t-d}) \left( b_0 + \sum_{j=1}^p b_j X_{t-j} \right) \quad (19)$$

where  $I(x)$  is a smooth function, typically a *distribution function*,  
ex.  $I(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

# GARCH model

A popular model in Economics and/or Finance is the GARCH model

$$X_t = \sigma_t Z_t \quad (20)$$

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (21)$$

where  $\mathbf{E}[Z] = 0$ ,  $\mathbf{Var}[Z] = 1$

Turns out that this is an ARMA model for  $X_t^2$ . Proof: Introduce a new white noise sequence  $\eta_t = X_t^2 - \sigma_t^2$

Many extensions based on SETAR or STAR!

# Bilinear models

The bilinear  $BL(p, q, m, k)$  model is defined as

$$X_t + \sum_{j=1}^p a_j X_{t-j} = \sum_{j=0}^q c_j e_{t-j} + \sum_{i=1}^m \sum_{j=1}^k b_{ij} X_{t-i} e_{t-j} \quad (22)$$

Autocorrelation is identical to a linear model, but the qualitative properties of the bilinear model is very different (the bilinear term can cause temporary "explosions")

# Random coefficient model

The random coefficient AR model (RCAR) is defined as

$$X_t = \sum_{i=1}^k (\beta_i + B_i(t)) X_{t-i} \quad (23)$$

where  $B_i(t)$  are *iid* random variables.

The stability of the RCAR models is generally worse than that of the corresponding AR models.

Stability analysis of RCAR models provides a mathematically sound approach to robustness checks - cf. Lyapunov theory.

# Comparison between models

We have simulated the model

$$X_n = 3.9X_{n-1}(1 - X_{n-1}) + U_n(X_{n-1}) \quad (24)$$

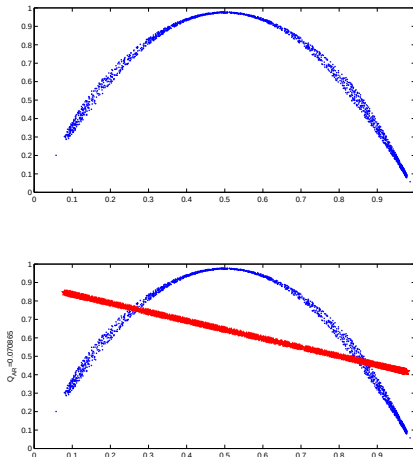
where  $U_n$  is constrained to prevent the process from leaving the space  $\chi = [0, 1]$ .

We used  $N_{Estimation} = 1\,000$  observations to estimate the models and  $N_{Eval} = 1\,000$  observations to evaluate the fitted models.

Quadratic loss out of sample comparison between models

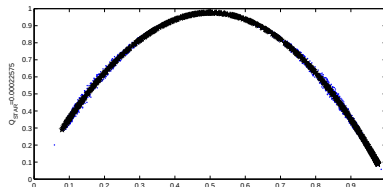
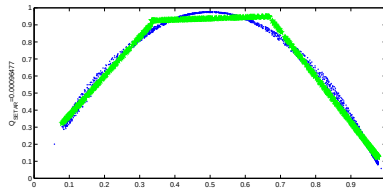
# Comparison between models

Data and fitted linear model,  $Q = 0.071$ .



# Comparison between models

SETAR (top) and STAR (below). The STAR model ( $Q = 0.00022$ ) can approximate this model rather well, compared to the linear and SETAR ( $Q = 0.00096$ ) models.



## Indep. Governed AR models (IGAR)

The regime in this class of models is determined by a stochastic variable  $J_t$ . The  $IGAR(l, d, (k_1), \dots, (k_l))$  model is defined as:

$$X_t = a_0^{(J_t)} + \sum_{i=1}^{k_{J_t}} a_i^{(J_t)} X_{t-i} + \epsilon_t^{(J_t)} \quad (25)$$

where

$$J_t = \begin{cases} 1 & \text{with prob } p_1 \\ 2 & \text{with prob } p_2 \\ \vdots & \vdots \\ l & \text{with prob } p_l \end{cases} \quad (26)$$

A particularly popular case appears when the regime variable  $J_t$  is given by a Markov chain. This is the  $MMAR(l, k_1, \dots, k_l)$  models

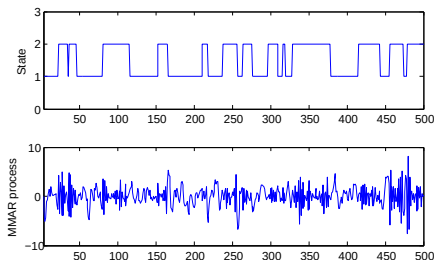
# MMAR(2,2,2) example

Two  $AR(2)$  processes

$$1 : X_n = 1.1X_{n-1} - 0.5X_{n-2} + e_n \quad (27)$$

$$2 : X_n = -1.2X_{n-1} - 0.5X_{n-2} + e_n \quad (28)$$

and  $P = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$ .



# Case study on HMMs

Time inhomogeneous Hidden Markov Models with applications in power systems

Lindström, E., Madsen, H., & Norén, V. (2015). Consumption management in the Nord Pool region: A stability analysis. *Applied Energy*, 146, 239-246.

<https://doi.org/10.1016/j.apenergy.2015.01.113>

# Jump models

A recent alternative to HMMs are Jump models

- ▶ Consider  $K$  possible states/regimes
- ▶ Parameters for each regime  $\theta_k$ ,  $k = 1, \dots, K$ .
- ▶ Penalty  $\lambda$  for switching between regimes  $s_{t-1} \rightarrow s_t$ ,  
 $t = 1, \dots, T$

Implied log-likelihood function (Bemporad et al, 2018) is given by

$$\mathcal{L}(\theta, \mathbf{s}, y, \mathbf{x}) = \sum_{t=1}^{T-1} [\ell(y_t, \theta_{s_t}) + \lambda \mathbb{I}_{\{s_t \neq s_{t-1}\}}] + \ell(y_T, \theta_{s_T}) \quad (29)$$

with  $\ell(y_t, \theta_{s_t}) = \|y_t - \mathbf{x}_t \theta_{s_t}\|_2^2$  (Nystrup et al, 2020, 2021)

# Jump models

A recent alternative to HMMs are Jump models

- ▶ Consider  $K$  possible states/regimes
- ▶ Parameters for each regime  $\theta_k$ ,  $k = 1, \dots, K$ .
- ▶ Penalty  $\lambda$  for switching between regimes  $s_{t-1} \rightarrow s_t$ ,  
 $t = 1, \dots, T$

Implied log-likelihood function (Bemporad et al, 2018) is given by

$$\mathcal{L}(\theta, \mathbf{s}, y, x) = \sum_{t=1}^{T-1} [\ell(y_t, \theta_{s_t}) + \lambda \mathbb{I}_{\{s_t \neq s_{t-1}\}}] + \ell(y_T, \theta_{s_T}) \quad (29)$$

with  $\ell(y_t, \theta_{s_t}) = \|y_t - x_t \theta_{s_t}\|_2^2$  (Nystrup et al, 2020, 2021)

# Difference between HMMs and Jump Models

- ▶ HMMs is soft classification of states, Jump models is hard classification, cf. Gaussian Mixture Models vs K-means(++)
- ▶ HMMs are difficult to estimate, log-likelihood function is severely multi-modal, in fact there are approximately as many local optima as observations (!)
- ▶ Jump models are estimated by alternating between estimating  $\theta|\mathbf{s}$  and  $\mathbf{s}|\theta$ . These sub-problems are convex!, essentially Viterbi algorithm and simple quadratic minimization.
- ▶ Exogenous variables and feature selection are possible in both frameworks, but scale much better for jump models (Nystrup et al, 2021).

# Difference between HMMs and Jump Models

- ▶ HMMs is soft classification of states, Jump models is hard classification, cf. Gaussian Mixture Models vs K-means(++)
- ▶ HMMs are difficult to estimate, log-likelihood function is severely multi-modal, in fact there are approximately as many local optima as observations (!)
- ▶ Jump models are estimated by alternating between estimating  $\theta|s$  and  $s|\theta$ . These sub-problems are convex!, essentially Viterbi algorithm and simple quadratic minimization.
- ▶ Exogenous variables and feature selection are possible in both frameworks, but scale much better for jump models (Nystrup et al, 2021).

# Difference between HMMs and Jump Models

- ▶ HMMs is soft classification of states, Jump models is hard classification, cf. Gaussian Mixture Models vs K-means(++)
- ▶ HMMs are difficult to estimate, log-likelihood function is severely multi-modal, in fact there are approximately as many local optima as observations (!)
- ▶ Jump models are estimated by alternating between estimating  $\theta|\mathbf{s}$  and  $\mathbf{s}|\theta$ . These sub-problems are convex!, essentially Viterbi algorithm and simple quadratic minimization.
- ▶ Exogenous variables and feature selection are possible in both frameworks, but scale much better for jump models (Nystrup et al, 2021).

# Difference between HMMs and Jump Models

- ▶ HMMs is soft classification of states, Jump models is hard classification, cf. Gaussian Mixture Models vs K-means(++)
- ▶ HMMs are difficult to estimate, log-likelihood function is severely multi-modal, in fact there are approximately as many local optima as observations (!)
- ▶ Jump models are estimated by alternating between estimating  $\theta|\mathbf{s}$  and  $\mathbf{s}|\theta$ . These sub-problems are convex!, essentially Viterbi algorithm and simple quadratic minimization.
- ▶ Exogenous variables and feature selection are possible in both frameworks, but scale much better for jump models (Nystrup et al, 2021).

# Difference between HMMs and Jump Models

- ▶ HMMs is soft classification of states, Jump models is hard classification, cf. Gaussian Mixture Models vs K-means(++)
- ▶ HMMs are difficult to estimate, log-likelihood function is severely multi-modal, in fact there are approximately as many local optima as observations (!)
- ▶ Jump models are estimated by alternating between estimating  $\theta|\mathbf{s}$  and  $\mathbf{s}|\theta$ . These sub-problems are convex!, essentially Viterbi algorithm and simple quadratic minimization.
- ▶ Exogenous variables and feature selection are possible in both frameworks, but scale much better for jump models (Nystrup et al, 2021).

# Case study on Cryptocurrencies using Sparse Jump Models

Comparison of a very large number of features, selecting those that drive cryptocurrency returns within a sparse statistical jump model framework

Cortese, F. P., Kolm, P. N., & Lindström, E. (2023). What drives cryptocurrency returns? A sparse statistical jump model approach. Digital Finance, 1-36.

<https://doi.org/10.1007/s42521-023-00085-x>

# Stability theory

Linear deterministic models:

- ▶ Continuous time

$$\frac{dy}{dt} = Ay \quad (30)$$

Stable iff  $\text{eig}(A) < 0$

- ▶ Discrete time

$$x_n = Ax_{n-1} \quad (31)$$

Stable iff  $|\text{roots}(A)| < 1$

# Stability theory

Linear deterministic models:

- ▶ Continuous time

$$\frac{dy}{dt} = Ay \quad (30)$$

Stable iff  $\text{eig}(A) < 0$

- ▶ Discrete time

$$x_n = Ax_{n-1} \quad (31)$$

Stable iff  $|\text{roots}(A)| < 1$

# Stability theory

Linear deterministic models:

- ▶ Continuous time

$$\frac{dy}{dt} = Ay \quad (30)$$

Stable iff  $\text{eig}(A) < 0$

- ▶ Discrete time

$$x_n = Ax_{n-1} \quad (31)$$

Stable iff  $|\text{roots}(A)| < 1$

# Stability theory

Linear deterministic models:

- ▶ Continuous time

$$\frac{dy}{dt} = Ay \quad (30)$$

Stable iff  $\text{eig}(A) < 0$

- ▶ Discrete time

$$x_n = Ax_{n-1} \quad (31)$$

Stable iff  $|\text{roots}(A)| < 1$

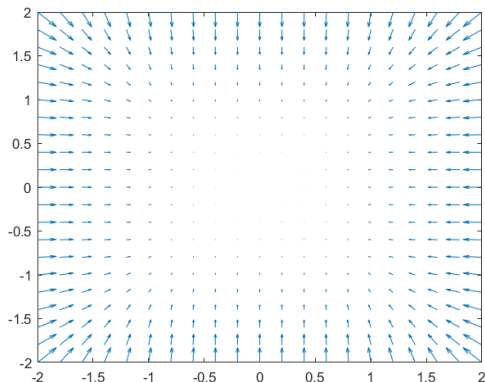
# What about non-linear models?

Continuous time model

$$\frac{dy_i}{dt} = -y_i^3, \quad i = 1, 2 \quad (32)$$

Can we say anything about this model?

# Vector field in 2 dimensions



Strong indications that  $(0,0)$  is a stable point.

# Different types of stability

Consider a system  $\dot{x}_t = f(x_t)$  where  $x_t \in \mathcal{D} \subseteq \mathbb{R}^d$  with equilibrium point  $f(x^*) = 0, x^* \in \mathcal{D}$ .

Three different stability definitions

- ▶ **Lyapunov stable** if there exist a  $\delta, \epsilon$  such that  $\|x(0) - x^*\| < \delta$  then  $\|x(t) - x^*\| < \epsilon$  for any  $\epsilon$  all  $t \geq 0$ .
- ▶ **Asymptotically stable** if *Lyapunov stable* and  $\lim_{t \rightarrow \infty} \|x_t - x^*\| = 0$ .
- ▶ **Exponentially stable** if *Asymptotically stable* and  $\|x(t) - x^*\| \leq \alpha \|x(0) - x^*\| e^{-\beta t}$  for  $\alpha, \beta > 0$ .

**Q:** What about the models discussed above?

# Different types of stability

Consider a system  $\dot{x}_t = f(x_t)$  where  $x_t \in \mathcal{D} \subseteq \mathbb{R}^d$  with equilibrium point  $f(x^*) = 0, x^* \in \mathcal{D}$ .

Three different stability definitions

- ▶ **Lyapunov stable** if there exist a  $\delta, \epsilon$  such that  $\|x(0) - x^*\| < \delta$  then  $\|x(t) - x^*\| < \epsilon$  for any  $\epsilon$  all  $t \geq 0$ .
- ▶ **Asymptotically stable** if *Lyapunov stable* and  $\lim_{t \rightarrow \infty} \|x_t - x^*\| = 0$ .
- ▶ **Exponentially stable** if *Asymptotically stable* and  $\|x(t) - x^*\| \leq \alpha \|x(0) - x^*\| e^{-\beta t}$  for  $\alpha, \beta > 0$ .

Q: What about the models discussed above?

# Different types of stability

Consider a system  $\dot{x}_t = f(x_t)$  where  $x_t \in \mathcal{D} \subseteq \mathbb{R}^d$  with equilibrium point  $f(x^*) = 0, x^* \in \mathcal{D}$ .

Three different stability definitions

- ▶ **Lyapunov stable** if there exist a  $\delta, \epsilon$  such that  $\|x(0) - x^*\| < \delta$  then  $\|x(t) - x^*\| < \epsilon$  for any  $\epsilon$  all  $t \geq 0$ .
- ▶ **Asymptotically stable** if *Lyapunov stable* and  $\lim_{t \rightarrow \infty} \|x_t - x^*\| = 0$ .
- ▶ **Exponentially stable** if *Asymptotically stable* and  $\|x(t) - x^*\| \leq \alpha \|x(0) - x^*\| e^{-\beta t}$  for  $\alpha, \beta > 0$ .

Q: What about the models discussed above?

# Different types of stability

Consider a system  $\dot{x}_t = f(x_t)$  where  $x_t \in \mathcal{D} \subseteq \mathbb{R}^d$  with equilibrium point  $f(x^*) = 0, x^* \in \mathcal{D}$ .

Three different stability definitions

- ▶ **Lyapunov stable** if there exist a  $\delta, \epsilon$  such that  $\|x(0) - x^*\| < \delta$  then  $\|x(t) - x^*\| < \epsilon$  for any  $\epsilon$  all  $t \geq 0$ .
- ▶ **Asymptotically stable** if *Lyapunov stable* and  $\lim_{t \rightarrow \infty} \|x_t - x^*\| = 0$ .
- ▶ **Exponentially stable** if *Asymptotically stable* and  $\|x(t) - x^*\| \leq \alpha \|x(0) - x^*\| e^{-\beta t}$  for  $\alpha, \beta > 0$ .

Q: What about the models discussed above?

# Different types of stability

Consider a system  $\dot{x}_t = f(x_t)$  where  $x_t \in \mathcal{D} \subseteq \mathbb{R}^d$  with equilibrium point  $f(x^*) = 0, x^* \in \mathcal{D}$ .

Three different stability definitions

- ▶ **Lyapunov stable** if there exist a  $\delta, \epsilon$  such that  $\|x(0) - x^*\| < \delta$  then  $\|x(t) - x^*\| < \epsilon$  for any  $\epsilon$  all  $t \geq 0$ .
- ▶ **Asymptotically stable** if *Lyapunov stable* and  $\lim_{t \rightarrow \infty} \|x_t - x^*\| = 0$ .
- ▶ **Exponentially stable** if *Asymptotically stable* and  $\|x(t) - x^*\| \leq \alpha \|x(0) - x^*\| e^{-\beta t}$  for  $\alpha, \beta > 0$ .

**Q:** What about the models discussed above?

# Generalized energy functions

Consider a non-linear system  $\dot{x} = f(x)$ , and function  $V(x) : \mathcal{D} \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$

$V$  is called a *Lyapunov candidate* if

- ▶  $V(x) = 0$  iff  $x = x^*$
- ▶  $V(x) > 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$
- ▶ Generalized energy is dissipating

cont:  $\dot{V}(x) = - \sum_{i=1}^d \frac{\partial V}{\partial x_i} \frac{\partial x_i}{\partial t} \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$

disc:  $V(x_{n+1}) - V(x_n) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$

Lyapunov theory allows us to establish stability without knowing trajectories.

# Generalized energy functions

Consider a non-linear system  $\dot{x} = f(x)$ , and function  $V(x) : \mathcal{D} \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$

$V$  is called a *Lyapunov candidate* if

- ▶  $V(x) = 0$  iff  $x = x^*$
- ▶  $V(x) > 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$
- ▶ Generalized energy is dissipating

**cont:**  $\dot{V}(x) = -\sum_{i=1}^d \frac{\partial V}{\partial x_i} \frac{\partial x_i}{\partial t} \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$

**disc:**  $V(x_{n+1}) - V(x_n) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$

Lyapunov theory allows us to establish stability without knowing trajectories.

## Two theorems

Assume that  $V$  is a Lyapunov candidate. Furthermore,

- ▶ If  $\dot{V}(x) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is Lyapunov stable.
- ▶ If  $\dot{V}(x) < 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is asymptotically stable.
- ▶ If  $\dot{V}(x) < -\beta V$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is exponentially stable with  $\|x(0) - x^*\| \leq M \|x(0) - x^*\| e^{-\beta t}$

"Proof": Note that  $V(x(t)) = V(x(0)) + \int_0^t \dot{V}(x(s)) ds \leq V(x(0))$

## Two theorems

Assume that  $V$  is a Lyapunov candidate. Furthermore,

- ▶ If  $\dot{V}(x) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is Lyapunov stable.
- ▶ If  $\dot{V}(x) < 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is asymptotically stable.
- ▶ If  $\dot{V}(x) < -\beta V$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is exponentially stable with  $\|x(t) - x^*\| \leq M\|x(0) - x^*\|e^{-\beta t}$

"Proof": Note that  $V(x(t)) = V(x(0)) + \int_0^t \dot{V}(x(s))ds \leq V(x(0))$

## Two theorems

Assume that  $V$  is a Lyapunov candidate. Furthermore,

- ▶ If  $\dot{V}(x) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is Lyapunov stable.
- ▶ If  $\dot{V}(x) < 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is asymptotically stable.
- ▶ If  $\dot{V}(x) < -\beta V$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is exponentially stable with  $\|x(t) - x^*\| \leq M\|x(0) - x^*\|e^{-\beta t}$

"Proof": Note that  $V(x(t)) = V(x(0)) + \int_0^t \dot{V}(x(s))ds \leq V(x(0))$

## Two theorems

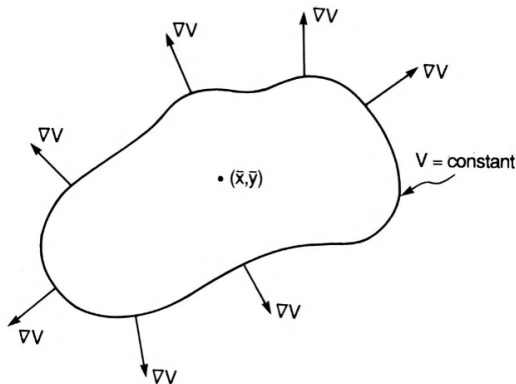
Assume that  $V$  is a Lyapunov candidate. Furthermore,

- ▶ If  $\dot{V}(x) \leq 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is Lyapunov stable.
- ▶ If  $\dot{V}(x) < 0$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is asymptotically stable.
- ▶ If  $\dot{V}(x) < -\beta V$  for all  $x \in \mathcal{D} \setminus \{x^*\}$ , then the system is exponentially stable with  $\|x(t) - x^*\| \leq M\|x(0) - x^*\|e^{-\beta t}$

"Proof": Note that  $V(x(t)) = V(x(0)) + \int_0^t \dot{V}(x(s))ds \leq V(x(0))$

# Level sets of $V$

$\dot{V}(x) < 0$  decreasing means that the level set is shrinking



# Stochastic stability

- ▶ Standard, deterministic theory can be difficult to apply to stochastic systems.
- ▶ Instead, consider
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] \leq 0$  for Lyapunov stability
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] < 0$  for asymptotical stability
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] < -\beta V(x_n)$  for exponential stability

# Stochastic stability

- ▶ Standard, deterministic theory can be difficult to apply to stochastic systems.
- ▶ Instead, consider
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] \leq 0$  for Lyapunov stability
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] < 0$  for asymptotical stability
  - ▶  $\mathbb{E}[V(x_{n+1}) - V(x_n)|\mathcal{F}_n] < -\beta V(x_n)$  for exponential stability

# Example 1

Consider the system

$$x_{n+1} = ax_n, \quad |a| < 1 \quad (33)$$

and the Lyapunov candidate function  $V(x) = x^2$

- ▶ Convince yourself that this is a valid Lyapunov function
- ▶ Write down the exact solution - what type of convergence do you expect?
- ▶ Show that this is in fact the case using Lyapunov theory

# Random Coefficient AR model

Extend the previous model by considering the model

$$x_{n+1} = (a + \xi)x_n, \quad |a| < 1 \quad (34)$$

with  $E[\xi] = 0$  and  $V[\xi] = \sigma^2$ .

- ▶ Find a suitable Lyapunov candidate function
- ▶ Try to find an upper limit for  $\sigma^2$  such that the system still is stable in some form.
- ▶ Extra: See if you can find a sharp limit a some special cases (e.g.  $a = 0$  and two point distribution or  $a \approx 1$ )

## Example 3

- ▶ Consider  $X_1, \dots, X_N \sim N(0, \sigma^2)$ . We know that  $\bar{X}_N \rightarrow 0$  according to LLN.
- ▶ Prove the stability (and type of stability using Lyapunov theory)
- ▶ Hint. Define  $S_N = \sum_{i=1}^N X_i$ . It then holds that

$$\bar{X}_N = \frac{1}{N} (X_N + S_{N-1}) \quad (35)$$

$$= \dots = \bar{X}_{N-1} + \frac{1}{N} (X_N - \bar{X}_{N-1}) \quad (36)$$

The rate of convergence is known from the central limit theory. Hence, do not expect to exceed that rate...