Advanced Econometrics

Chapter 3: Probability Models

Etienne Wijler

Econometrics and Data Science
Master Program

VRIJE UNIVERSITEIT AMSTERDAM
SCHOOL OF BUSINESS AND ECONOMICS

What we did until now!

Chapters 1 and 2:

1. Reviewed simple linear models!
2. Identified a problem:
   - These are not the models you will implement in your work!
     - Private companies, central banks, governments, public institutions, research institutions, universities
   - Simple linear models everyone can estimate! (click buttons) Econometricians (are expected to) do more!
   - Econometricians are suppose to be specialists in cutting-edge, state-of-the-art models (not available in software packages).

Chapter 3: Probability Models

Chapter 3: Probability models

> **1st part:** many new definitions and concepts
>
> **2nd part:** state-of-the-art nonlinear dynamic probability models

Need a RECAP of basic probability and statistics?

Then read Appendix A!

- Probability spaces
- Random variable
- σ-algebras and measurable spaces

Chapter 3: Probability models

Some more reading material:

1. Davidson (1994), *"Stochastic Limit Theory"*
   - Chapter 1.6, 2.3, 3.1 and 7.1
2. Billingsley (1995), *"Probability and Measure"*
   - Chapter 2 and 5
3. White (1996), *"Estimation Inference and Specification Analysis"*
   - Chapter 2.1, 2.2 and 20
4. Fan and Yao (2005), *"Nonlinear Time-Series"*
   - Chapter 1.3

What is a probability model?

Question: What exactly is a model?

Example: Given T tosses of a coin, it is reasonable to suppose that $x_1, ..., x_T$ are realizations of T Bernoulli random variables, $x_t \sim \text{Bern}(\theta)$ with unknown probability parameter $\theta \in [0, 1]$.

Important: Each θ defines a probability distribution for the random vector $(x_1, ..., x_T)$ taking values in $\mathbb{R}^T$. Our model is a *collection of probability distributions* on $\mathbb{R}^T$.

This definition of model is the one you have been always using. Even if you did not realize it!

What is a probability model?

Question: What exactly is a model?

Example 2: Gaussian linear AR(1) model,

$$x_t = \alpha + \beta x_{t-1} + \varepsilon_t \ , \quad \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \ , \qquad \forall \ t \in \mathbb{Z}$$

Important: Each $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2_\varepsilon)$ defines a distribution for the time-series $\{x_t\}_{t\in\mathbb{Z}}$. Our model is a *collection of probability distributions* on $\mathbb{R}^\infty$.

This definition of model is the one you have been always using. Even if you did not realize it!

Probability model

Definition (Probability model)

Given a parameter space Θ, a probability model is a collection $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ of probability measures defined by each $\boldsymbol{\theta} \in \Theta$.

Useful definitions...

Definition ((Non)Parametric models)

A probability model $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ is said to be:

- 'parametric' if the parameter space Θ is finite dimensional;
- 'nonparametric' if Θ is infinite dimensional;
- 'semi-parametric' if $\Theta = \Theta_1 \times \Theta_2$ where Θ_1 is finite dimensional and Θ_2 is infinite dimensional;
- 'semi-nonparametric' if Θ_T is indexed by the sample size T with 'sieves' $\{\Theta_T\}_{T \in \mathbb{N}}$ with increasing dimension.

Definition (Nested models)

Given two parametric models $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ and $\mathbb{P}^*_{\Theta^*} := \{P^*_{\boldsymbol{\theta}^*}, \boldsymbol{\theta}^* \in \Theta^*\}$, we say that model $\mathbb{P}_\Theta$ nests model $\mathbb{P}^*_{\Theta^*}$ if and only if $\mathbb{P}^*_{\Theta^*} \subseteq \mathbb{P}_\Theta$.

Probability model

Question: *why should we work with probability models?*

A very brief history of econometrics:

1. In 1936, Jan Tinbergen published macroeconometric model of The Netherlands with hundreds of regressions.

   > This *'was a major force in the transformation of economics from a discursive discipline into a model-building discipline'. (Solow)*

2. In 1939, John Maynard Keynes published critical review of Tinbergen's work: *the model could not be proven wrong!* Error term accounts for anything!

3. In 1944, Haavelmo published *'The Probability Approach in Econometrics'* which solved the problem: unlikely errors consitute evidence against the model!

What is a DGP?

Definition (DGP)

The data generating process of a random variable x is the probability measure P_0 that defines the stochastic behavior of x.

Definition (DGP of a time series)

The data generating process of a time series $\{x_t\}_{t\in\mathbb{Z}}$, is the probability measure P_0 that defines the stochastic behavior of the infinite random sequence.

Intuition: The DGP is the 'unknown mechanism' that 'generates the data'.

Note: In economics, the DGP is most likely very complex involving millions of agents, factors, variables, decisions, etc.

Correct Specification

Definition (Correct specification)

A model $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ is said to be correctly specified if the data generating process P_0 is an element of the model $\mathbb{P}_\Theta$; i.e. if there exists $\boldsymbol{\theta}_0 \in \Theta$ such that $P_{\boldsymbol{\theta}_0} = P_0$. When this parameter $\boldsymbol{\theta}_0$ exists, then it is called the 'true parameter'.

Shorter notation: a model is correctly specified if $\exists\ \boldsymbol{\theta}_0 \in \Theta : P_{\boldsymbol{\theta}_0} = P_0$ which means that $P_0 \in \mathbb{P}_\Theta$.

Definition (Mis-specification)

A model $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ is said to be mis-specified (or incorrectly specified) if the data generating process P_0 is not an element of the model $\mathbb{P}_\Theta$; i.e. if $P_{\boldsymbol{\theta}} \neq P_0\ \forall\ \boldsymbol{\theta} \in \Theta$.

Correct specification

Example: (`Linear Gaussian AR(1)model`)
For every $\boldsymbol{\theta} := (\alpha, \beta, \sigma_\varepsilon^2) \in \Theta \subseteq \mathbb{R}^3$ let

$$x_t = \alpha + \beta x_{t-1} + \varepsilon_t \ , \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \ , \qquad \forall\, t \in \mathbb{Z}.$$

Note: This is a model because $\boldsymbol{\theta}$ can take many values in Θ

Note: The collection of probability measures on the adopted σ-algebra of $\mathbb{R}^\infty$ is specified implicitly.

Specification: Given a time series $\{x_t\}_{t\in\mathbb{Z}}$ with probability measure P_0, the AR(1) model is well specified if $\exists\ \boldsymbol{\theta}_0 \in \Theta$ such that the linear AR(1) generates a random sequence with measure $P_{\boldsymbol{\theta}_0} = P_0$.

Stationarity and white noise

Definition (Weak stationarity)

A time series $\{x_t\}$ is weakly stationary if $\mu_t = \mathbb{E}(x_t)$ and $\gamma_t(h) = \mathbb{C}\text{ov}(x_t, x_{t-h})$ satisfy $\mu_t = \mu$ and $\gamma_t(h) = \gamma(h)\ \forall\ (t, h)$.

In words: the (unconditional) mean, variance and covariances are *constant* over time.

Note: weak stationarity is sometimes also referred to as *second-order stationarity*.

Definition (White noise process)

A random sequence $\{x_t\}_{t\in\mathbb{Z}}$ is said to be a white noise process, if $\mathbb{C}\text{ov}(x_t, x_{t-h}) = 0$, $\mathbb{E}(x_t) = 0$ and $\mathbb{V}\text{ar}(x_t) = \sigma^2\ \forall\ (t, h)$.

Linear model

Definition (Linear time series)

A time-series $\{x_t\}$ is said to be linear if it can be represented as,

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j z_{t-j} \quad \text{where } \{z_t\} \sim \text{WN}(0, \sigma^2)$$

and $\{\psi_j\}_{-\infty}^{\infty}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Definition (Linear time series model)

A time-series model $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ is said to be linear if every measure $P_{\boldsymbol{\theta}}$ defines a stochastic process that is linear.

Wold's representation theorem

Theorem (Wold's representation theorem)

Let $\{x_t\}$ be a weakly stationary process. Then it admits the following representation,

$$x_t = \sum_{j=0}^{\infty} \psi_j z_{t-j} + v_t, \qquad \textit{where}$$

a. $\psi_0 = 1$ *and* $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,

b. $\{z_t\}_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$

c. $\{v_t\}_{t\in\mathbb{Z}}$ *is deterministic (non-random).*

Herman Wold was a famous Norwegian-born statistician whom you may know from the Cramér-Wold theorem, an essential tool in deriving the joint convergence of multiple sequences of random variables.

Introductory econometrics: Wold's theorem was used as a justification for the adoption of linear dynamic models.

Example: (`ARMA(`p,q`) model`)

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim$NID$(0, \sigma^2_\varepsilon)$.

Recall: ARMA(p,q) is linear if the autoregressive polynomial $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ is invertible, because then it can be re-written in the infinite MA representation

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with } \psi_0 = 1 \text{ and } \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Problems with this justification...

Wold's representation $\quad + \quad$ MA(∞) Representation of ARMA

Seems that ARMA models can describe any WS process!

This is not true! We must look at the details:

1. Stochastic component of Wold's theorem $\sum_{j=0}^{\infty} \psi_j z_{t-j}$ is not necessarily a linear process.
2. Stochastic component of Wold's theorem $\sum_{j=0}^{\infty} \psi_j z_{t-j}$ involves infinitely many parameters $\{\psi_j\}$ that cannot be estimated from a finite sample of data.
3. Wold's theorem features a deterministic component $\{v_t\}_{t \in \mathbb{Z}}$ that is unknown and potentially very complex.
4. Distribution of the white noise sequence $\{z_t\}$ is unknown and possibly very complex!

Nonlinear dynamic models

Definition (Nonlinear time series models)

A time-series model $\mathbb{P}_\Theta := \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ is said to be nonlinear if at least some measure $P_{\boldsymbol{\theta}} \in \mathbb{P}_\Theta$ defines a stochastic process that is not linear.

Example: (`Quadratic AR model`)

$$x_t = \alpha + \beta x_{t-1} + \gamma x_{t-1}^2 + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

is *nonlinear model* if Θ allows for $\gamma \neq 0$.

Example: (`SESTAR model`)

$$x_t = \gamma/(1 + \exp(\alpha + \beta x_{t-1}))x_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

is *nonlinear model* if $\gamma \neq 0$ and $\beta \neq 0$.

Note: Both models above nest the linear AR(1) model.

Examples of nonlinear dynamic models

Some more reading material:

1. *Fan and Yao (2005), "Nonlinear Time-Series"*
   - *Chapter 1.4, 1.5 and 4*
2. *Granger and Terasvirta (1993), "Modeling Nonlinear Economic Relationships"*
   - *Chapter 2.1, 3.1 and 7.3*

Nonlinear autoregressions: GNLAR

Let $\boldsymbol{\theta} \in \Theta$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be (iid) innovations with $\varepsilon_t \sim p_\varepsilon(\boldsymbol{\theta})$.

Example: (`GNLAR(1) model`) The *general nonlinear autoregressive* (GNLAR) model is given by,

$$x_t = f(x_{t-1}, \varepsilon_t; \boldsymbol{\theta}) \quad \forall\, t \in \mathbb{Z}$$

Example: (`GNLAR(`p`) model`) The GNLAR(p) takes the form

$$x_t = f(x_{t-1}, ..., x_{t-p}, \varepsilon_t; \boldsymbol{\theta}) \quad \forall\, t \in \mathbb{Z}.$$

Example: (`GNLARMA(`p, q`) model`) *general nonlinear autoregressive moving-average* GNLARMA(p,q) model takes the form

$$x_t = f(x_{t-1}, ..., x_{t-p}, \varepsilon_t, ..., \varepsilon_{t-q}; \boldsymbol{\theta}) \quad \forall\, t \in \mathbb{Z}.$$

Nonlinear autoregressions: NLAR

Example: (`NLAR(1) model`) the *nonlinear autoregressive* (NLAR) model with additive innovations

$$x_t = f(x_{t-1}; \boldsymbol{\theta}) + \varepsilon_t \quad \forall\, t \in \mathbb{Z}.$$

Note: NLAR(1) models are often written in the equivalent form

$$x_t = g(x_{t-1}; \boldsymbol{\theta}) x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

by defining $g(x_{t-1}; \boldsymbol{\theta}) := f(x_{t-1}; \boldsymbol{\theta})/x_{t-1}$, or

$$x_t = \beta_t x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

by defining $\beta_t := g(x_{t-1}; \boldsymbol{\theta})$.

Nonlinear autoregressions: TAR Model

Example: (`threshold autoregressive (TAR) model`) is given by,

$$x_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + \varepsilon_t & \text{if} \;\; z_{t-1} \leq r \\ \alpha_2 + \beta_2 x_{t-1} + \varepsilon_t & \text{if} \;\; z_{t-1} > r \end{cases} \qquad \forall \; t \in \mathbb{Z}$$

Note: TAR allows for different *dynamic regimes*

Note: TAR can also be written as

$$x_t = (\alpha_1 + \beta_1 x_{t-1})\mathbf{1}_{(-\infty < z_{t-1} \leq r)} + (\alpha_2 + \beta_2 x_{t-1})\mathbf{1}_{(r < z_{t-1} < \infty)} + \varepsilon_t,$$

Nonlinear autoregressions: TAR model

***Exogenous TAR* model** when the driver $\{z_t\}$ is exogenous

***Self-excited TAR* (SETAR):** when driver is the lagged dependent variable $\{x_{t-1}\}$

$$x_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + \varepsilon_t & \text{if } \; x_{t-1} \leq r \\ \alpha_2 + \beta_2 x_{t-1} + \varepsilon_t & \text{if } \; x_{t-1} > r \end{cases} \qquad \forall\, t \in \mathbb{Z}.$$

Generality: the TAR can accomodate k different regimes,

$$x_t = \sum_{i=1}^{k} (\alpha_i + \beta_i x_{t-1}) \mathbf{1}_{r_{i-1} \leq z_{t-1} < r_i} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

where $-\infty = r_0 < r_1 < ... < r_k = \infty$ are the regimes boundaries

Nonlinear autoregressions: STAR model

Example: (`STAR model`) A famous NLAR(1) model is the *smooth transition autoregressive* (STAR) model,

$$x_t = g(z_{t-1};\boldsymbol{\theta})x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

where $\{\varepsilon_t\}$ are innovations with some specified distribution and

$$g(z_{t-1};\boldsymbol{\theta}) := \delta + \frac{\gamma}{1+\exp(\alpha+\beta z_{t-1})} \quad \forall\, t \in \mathbb{Z}.$$

Alternative STAR models:

1. *exogenous STAR* model (exogenous driver $\{z_t\}$)
2. *logistic self-excited STAR* (endogenous driver $z_t = x_t$)
3. *exponential self-excited* STAR (endogenous driver $z_t = x_t^2$)

Logistic SESTAR

Example: (`Logistic SESTAR model`)

$$g(z_{t-1};\boldsymbol{\theta}) := \delta + \frac{\gamma}{1+\exp(\alpha+\beta x_{t-1})},\ \forall\, t \in \mathbb{Z}.$$

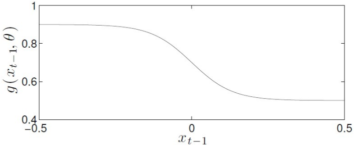


Figure: Plot of $g(x_{t-1};\boldsymbol{\theta})$ for $(\delta,\gamma,\alpha,\beta) = (0.5, 0.4, 0, 15)$.

Note: the logistic SESTAR model allows us to model changes in the dependence of the time-series $\{x_{t-1}\}_{t\in\mathbb{Z}}$.

Practice: crucial for modeling higher dependence of macro variables in recessions $(x_{t-1} < 0)$ than expansions $(x_{t-1} > 0)$

Exponential SESTAR

Example: (`Exponential SESTAR model`)

$$g(z_{t-1};\boldsymbol{\theta}) := \delta + \frac{\gamma}{1+\exp(\alpha+\beta(x_{t-1}-\mu)^2)} \qquad \forall\, t \in \mathbb{Z}.$$

Practice: modelling real foreign exchange (FX) rate: high-dependence (no mean reversion) at rates near 1 and low-dependence (mean reverting behavior) at rates far from 1.

Note: behavior of real FX is justified by the *law of one price*!

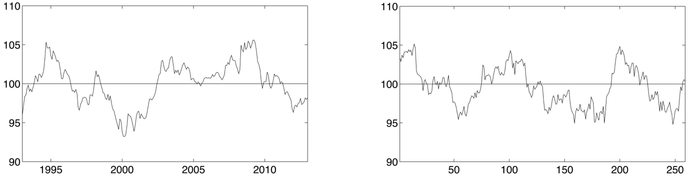


Figure: FX of EUR/DK (left) vs data simulated by Exponentional SESTAR (right).

Distributed SESTAR

Example: (`Distributed SESTAR`)

$$x_t = a_t x_{t-1} + b_t w_t + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

$$a_t := \delta + \frac{\gamma}{1 + \exp(\alpha + \beta x_{t-1})} \quad , \quad b_t := \delta^* + \frac{\gamma}{1 + \exp(\alpha^* + \beta^* x_{t-1})}.$$

Note: this model generates time-varying multiplier effect!

Practice: crucial extension to ADL model for explaining changes in government expenditure multiplier!

Economic theory: predicts that multiplier is larger when economy is substantially below potential.

Economic policy: macro $\neq$ micro (not like a family!)

Random coefficient autoregression

Example: (`RCAR(1) model`) the *Random coefficient autoregressive* (RCAR) model is given by

$$x_t = \phi_{t-1} x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

where both $\{\phi_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are exogenous iid sequences with a certain distribution.

Note: first proposed by Quinn (1980), it has several important applications in finance and biology as it allows for a time-varying conditional mean and variance.

$$\phi_{t-1} \sim N(\phi, \sigma_\phi^2) \quad \text{and} \quad \varepsilon_t \sim N(0, \sigma_\epsilon^2) \quad \forall\, t \in \mathbb{Z}.$$

$$x_t | x_{t-1} \ \sim \ N\Big(\phi x_{t-1} \ , \ \ \sigma_\phi^2 x_{t-1}^2 + \sigma_\varepsilon^2\Big).$$

Two classes of time-varying parameter models

Recall: the NLAR(1) model

$$x_t = f(x_{t-1}; \boldsymbol{\theta}) + \varepsilon_t$$

can be re-written as a time-varying parameter model

$$x_t = \phi_{t-1} x_{t-1} + \varepsilon_t$$

by defining ϕ_{t-1} as $\phi_{t-1} := f(x_{t-1}; \boldsymbol{\theta})/x_{t-1}$.

NLAR: ϕ_t is endogenous! **RCAR:** ϕ_t is exogenous!

Important: Difference between NLAR(1) and RCAR(1) is the driver of the time-varying parameter $\{\phi_t\}$!

Cox (1981): Two classes of *time-varying parameter models*

1. *observation-driven* (endogenous)
2. *parameter-driven* (exogenous)

Local level model: parameter-driven

Example: `(Gaussian local-level model)` The time-varying mean μ_t takes the form,

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{NID}(0, \sigma^2_\varepsilon) \ ,$$

$$\mu_t = \omega + \beta \mu_{t-1} + v_t \ , \quad \{v_t\} \sim \text{NID}(0, \sigma^2_v).$$

Note: μ_t evolves in time independently of $\{x_t\}_{t \in \mathbb{Z}}$. This is what makes the model a *parameter-driven model*!

Practice: $\{\mu_t\}$ can be filtered using the Kalman filter.

Local-level model: observation-driven

Example: (`Gaussian Local-level model`)

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \mathrm{NID}(0, \sigma_\varepsilon^2) \ ,$$

$$\mu_t = \omega + \alpha(x_{t-1} - \mu_{t-1}) + \beta\mu_{t-1}.$$

Note: the update of μ_t is determined by the lagged observation x_{t-1} plus an autoregressive term. This makes the model an *observation-driven model.*

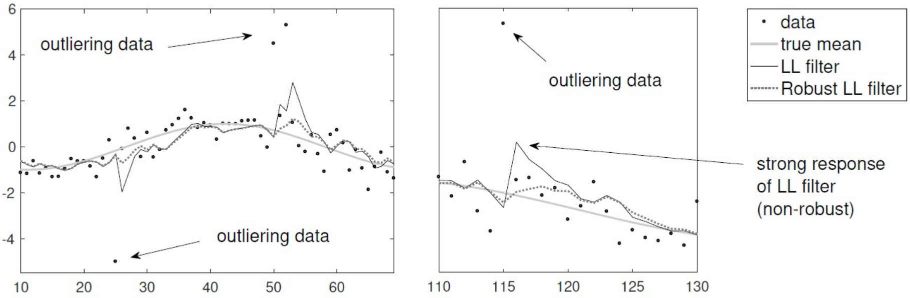
Important: When the signal-to-noise ratio is small, it may be difficult to extract the path of the time-varying mean. In the presence of outliers, the filter of μ_t may fluctuate too much. <u>Robust filters are important</u>.

Robust Local-level model: observation-driven

Example: (Robust local-level model)

$$x_t = \mu_t + \varepsilon_t \;, \quad \{\varepsilon_t\} \sim \text{iid } \tau(\lambda) \;,$$

$$\mu_t = \omega + \alpha \tanh(x_{t-1} - \mu_{t-1}) + \beta\mu_{t-1} \;,$$



Volatility model: parameter driven

Example: (Gaussian `stochastic volatility model`)

$$x_t = \sigma_t \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \mathrm{NID}(0, \sigma_\varepsilon^2) \ ,$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + v_t \ , \quad \{v_t\} \sim \mathrm{NID}(0, \sigma_v^2).$$

Note: σ_t^2 evolves in time independently of $\{x_t\}_{t \in \mathbb{Z}}$. This is what makes the model a *parameter-driven model*!

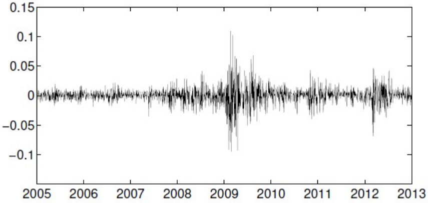
Practice: $\{\sigma_t^2\}$ can be filtered using the Kalman filter.

Practice: Many applications in finance!

Volatility model: observation-driven

Example: (`GARCH model`) the *generalized autoregressive heteroeskedasticity* model of Engle (1982) and Bollerslev (1986)

$$x_t = \sigma_t \varepsilon_t \ , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0,1),$$
$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2.$$



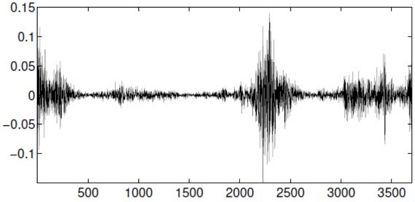


Figure: S&P500 returns (left). Simulated data (right).

Note: GARCH filter is too sensitive to outliers (not robust).

Robust GARCH

Example: (`Robust GARCH model`)

$$x_t = \sigma_t \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{iid } \tau(\lambda) \ ,$$

$$\sigma_t^2 = \omega + \alpha \frac{x_{t-1}^2}{1 + x_{t-1}^2} + \beta \sigma_{t-1}^2 .$$

Note: The volatility *update* is uniformly bounded x_{t-1}.

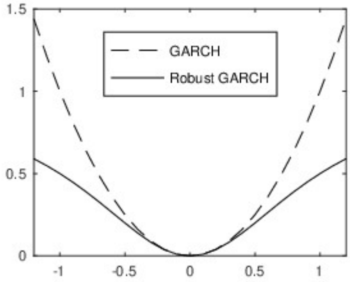
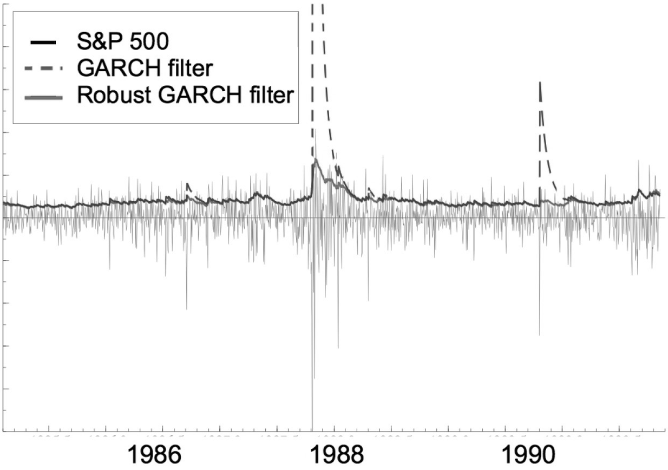


Figure: *News impact curve* of GARCH and Robust GARCH models.

GARCH vs Robust GARCH



Leverage effect: NGARCH and QGARCH

Leverage effect: Negative shocks produce more volatility that positive shocks!

Example: (`Q-GARCH model`) the *quadratic GARCH* (QGARCH) of Sentana (1995)

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \gamma x_{t-1} + \beta \sigma_{t-1}^2.$$

Example: (`NGARCH model`) the *nonlinear GARCH* (NGARCH) proposed by Engle and Ng (1993) takes the form

$$\sigma_t^2 = \omega + \alpha (x_{t-1} - \delta \sigma_{t-1})^2 + \beta \sigma_{t-1}^2.$$

Note: $\alpha (x_{t-1} - \delta \sigma_{t-1})^2 = \alpha x_{t-1}^2 + \alpha \delta^2 \sigma_{t-1}^2 - \alpha \delta x_{t-1} \sigma_{t-1}.$

Dynamic factor models (DFMs):

$$x_{i,t} = \; g(\; f_t \; , \; \mu_{i,t} \; , \; \varepsilon_{i,t} \;)$$

- i=1,...,N and t = 1,...,T
- Decompose the dynamics of multiple time-series into:
  - a *common factor* f_t
  - an *idiosyncratic component* $\mu_{i,t}$
  - an *error term* $\varepsilon_{i,t}$
- Offer a simple and parsimonious way of incorporating information from several variables into a dynamic model

DFMs for the conditional mean

Example: (`Linear DFM for common conditional mean`)

$$x_{i,t} = f_t + \mu_{i,t} + \varepsilon_{i,t} \ ,$$

$$f_t = \kappa + \delta(\frac{1}{N}\sum_{j=1}^{N} x_{j,t-1} - f_{t-1}) + \gamma f_{t-1} \ ,$$

$$\mu_{i,t} = \omega_i + \alpha_i(x_{i,t-1} - f_{t-1} - \mu_{i,t-1}) + \beta_i \mu_{i,t-1} \ ,$$

Example: (`Robust DFM for common conditional mean`)

$$x_{i,t} = f_t + \mu_{i,t} + \varepsilon_{i,t} \ ,$$

$$f_t = \kappa + \delta(\frac{1}{N}\sum_{j=1}^{N} x_{j,t-1} - f_{t-1}) + \gamma f_{t-1} \ ,$$

$$\mu_{i,t} = \omega_i + \alpha_i \tanh(x_{i,t-1} - f_{t-1} - \mu_{i,t-1}) + \beta_i \mu_{i,t-1} \ ,$$

Application: European IPI

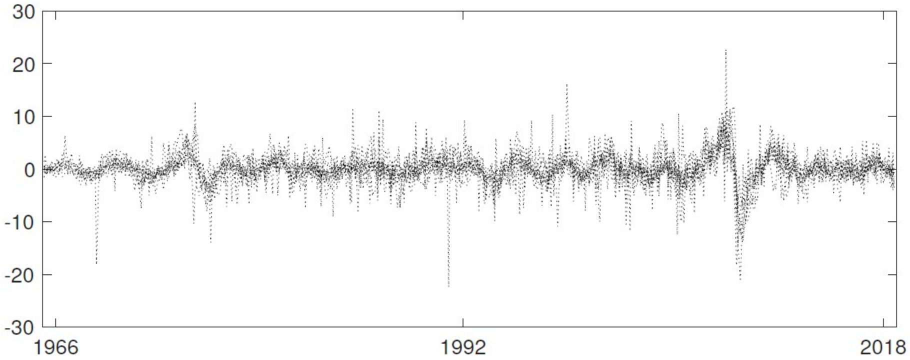


Figure: Plots of the Industrial Production Index (IPI) business cycle (HP-filtered data) for Euro area countries. Data shows a clear common factor (common variations in the IPI cycle) as well as country-specific idiosyncratic shocks and dynamics.

Application: European IPI

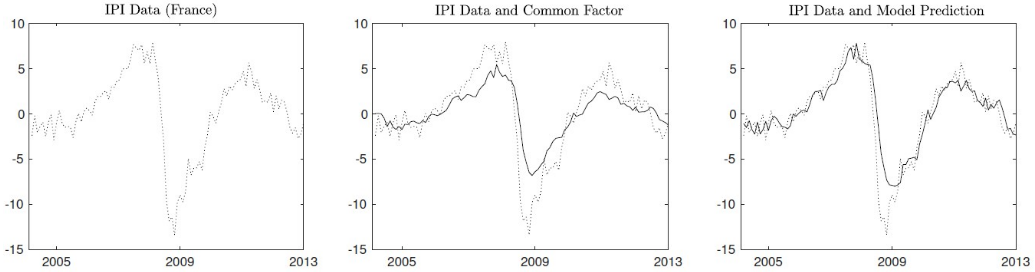


Figure: Decomposition of French IPI business cycle data (left) into a common Euro area factor (center), and a final prediction (right) which includes a common factor and an idiosyncratic component $\hat{x}_t = \hat{f}_t + \hat{\mu}_t$.

DFM: conditional volatility

Example: (`DFM for common conditional volatilities`) A nonlinear dynamic factor volatility model for mean zero time-series with a robust filter for the idiosyncratic volatilities can be defined as,

$$x_{i,t} = f_t \sigma_{i,t} \varepsilon_{i,t} \ ,$$

$$f_t^2 = \kappa + \delta\left(\frac{1}{N}\sum_{j=1}^{N} x_{j,t-1}^2 - f_{t-1}^2\right) + \gamma f_{t-1}^2 \ ,$$

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \tanh(x_{i,t-1}^2 - f_{t-1}^2 - \sigma_{i,t-1}^2) + \beta_i \sigma_{i,t-1}^2 \ ,$$

Semi-nonparametric models

Semi-nonparametric (SNP) models: are used to describe the relation between random variables in a very flexible way.

Example: (`Semi-nonparametric model`) An SNP model is a flexible model that can take the form

$$x_{t+1} = \sum_{i=1}^{k} \theta_i f_i(x_{t-1}; z_t, \boldsymbol{\theta}, \varepsilon_t) \quad \forall\, t \in \mathbb{Z}. \tag{1}$$

$f_1, f_2, ..., f_k$ are *basis functions* and $\theta_1, \theta_2, ..., \theta_k$ are parameters.

Note: the SNP model is designed to approximate complex unknown functions by setting a large $k \to \infty$ as $T \to \infty$

Semi-nonparametric models

Example: (`Autoregressive polynomial SNP model`) The autoregressive polynomial SNP model of order k, with additive error, takes the form

$$x_{t+1} = \sum_{i=0}^{k} \theta_i x_t^i + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

Stone–Weierstrass Theorem: we can approximate arbitrarily well the model

$$x_{t+1} = f_0(x_t) + \varepsilon_t$$

with polynomials of x_t.

Note: different choices of basis functions lead to different SNP models!! **Example:** Artificial Neural Networks

Artificial Neural Networks (ANN)

Example: (`Recurrent ANN`) A recurrent ANN with a single hidden-layer, k-neurons, and an exogenous input z_t, is given by

$$x_{t+1} = \sum_{i=0}^{k} a_i g_i(x_t, z_t; \boldsymbol{\theta}) + \varepsilon_t \quad \forall\, t \in \mathbb{Z} \ ,$$

$$\text{with} \quad g_i(x_t, z_t; \boldsymbol{\theta}) := \frac{1}{1 + \exp(w_{i,1} x_t + w_{i,2} z_t - b_i)} \ ,$$

- the parameters $w_{i,1}$ and $w_{i,2}$ are called the *input weights*,
- each $g_i(x_t, z_t; \boldsymbol{\theta})$ is called a neuron,
- Neurons in the first layer are called *inputs* (x_t and z_t),
- The 3rd layer, with the prediction $\hat{x}_{t+1}$, is called the *output*,
- b_is are the input *bias*, and a_is are *output weights*,
- function g_i is the *activation function*,
- The 2nd layer, composed of $g_i(x_t, z_t; \boldsymbol{\theta})$s is the *hidden layer*,
- This ANN is *recurrent* since the lag x_t is used as an input.

Artificial Neural Networks

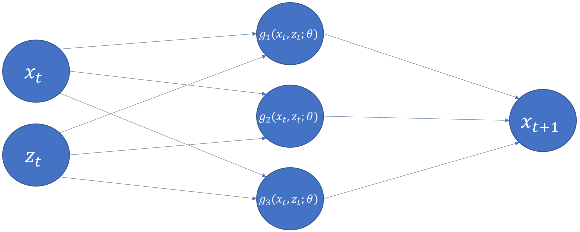


Figure: Graphical representation of an artificial neural network with two inputs, a single hidden layer with three neurons, and a single output.

The name ANN is used as it resembles, at an abstract level, the workings of a actual biological brain.

ANNs are very popular in machine learning and data science to train machines and understand large unstructured data sets

Deep Learning

In big data settings: the complexity and flexibility of ANNs can increase considerably by allowing for multiple hidden layers

Additional hidden layers: give us multiple transformations of the inputs

Deep learning: the use of ANNs with multiple hidden layers

Example: (`Multiple hidden-layer feedfoward ANN`) A feedforward ANN with two hidden-layers, featuring k_1 and k_2 neurons in each layer, is given by

$$x_{t+1} = \sum_{i=0}^{k_2} a_{i,2} g_{i,2}\big(\mathbf{g}_1(\mathbf{z}_t;\boldsymbol{\theta});\boldsymbol{\theta}\big) + \varepsilon_t \quad \forall\, t \in \mathbb{Z}\ ,$$

where $\mathbf{g}_1(\mathbf{z}_t;\boldsymbol{\theta})$ is a vector of k_1 neurons in the first layer

$$g_{i,1}(\mathbf{z}_t;\boldsymbol{\theta}) := \frac{1}{1+\exp(\sum_{j=1}^{n} w_{j,1}\mathbf{z}_{j,t} - b_{i,1})}$$

and each of the k_2 neurons in the second layer is given by

$$g_{i,2}\big(\mathbf{g}_1(\mathbf{z}_t;\boldsymbol{\theta});\boldsymbol{\theta}\big) := \frac{1}{1+\exp(\sum_{j=1}^{n} w_{j,2} g_{j,1}(\mathbf{z}_t;\boldsymbol{\theta}) - b_{j,2})}$$

Artificial Neural Networks

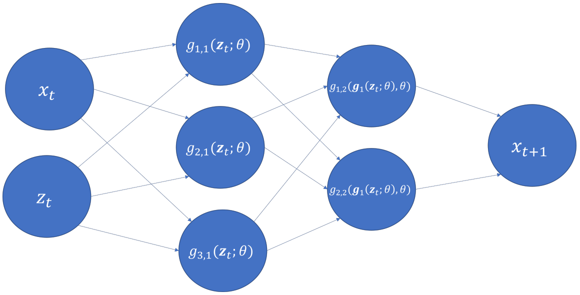


Figure: Graphical representation of an artificial neural network with two inputs, two hidden layers with three and two neurons respectively, and a single output.

Note: The mathematical and statistical properties were established by three econometricians! Hornik, Stinchcombe and White (1989) proved that multilayer feedforward ANNs are *universal approximators* of any measurable function. White (1989) shows that ANNs can be consistently *trained* (estimated), proving the validity of using ANNs for data analysis.

SNP models

In practice: SNP models have been successfully used in:

(a) handwriting and speech recognition;

(b) object recognition and navigation for self-driving vehicles;

(c) house pricing with detailed geographical features;

(d) classification of documents, images and videos available on the internet for improved search engine experience;

(e) identification of consumer groups for targeted marketing;

(f) creating machines with artificial intelligence.

However:

(a) tested in economics/finance for many decades: most developments date back to 60's, 70's and 80's

(b) simply not useful in small/medium sized samples

(c) most of econometrics is devoted to structural modeling, not predictive models (Chapter 11)