# ECON 616: Lecture 10: Intro to Nonlinear Times Series Models

Ed Herbst

# References

Books:

- Hamilton, Chapter 21

# Introduction

This week, we'll focus on a particular kind of nonlinearity – time-varying volatiltity.

1. ARCH/GARCH models
2. Stochastic Volatility Models
3. Markov-Switching Models

# ARCH

We started the course talking about *autoregressive models* for observables $y_t$. Let's allow the variance of $y_t$ to vary over time.

The p-th order ARCH model, first used in cite:Engle$_{1982}$

$$y_t = \mu + \epsilon_t \tag{1}$$
$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 \tag{2}$$
$$\epsilon_t = \sigma_t e_t \quad e_t \sim N(0,1). \tag{3}$$

Key feature: variance of $\epsilon_t$ is time varying and depends on past $p$ shocks through their squares.

Note: $\sigma_t$ is known at time $t-1$!

## ARCH

This means that

$$y_{t|t-1} \sim N(\mu, \sigma_t^2)$$

Since $E_{t-1}[\epsilon_t^2] = E_{t-1}[e_t^2 \sigma_t^2] = \sigma_t^2 E_{t-1}[e_t^2] = \sigma_t^2$.

What about the unconditional expectation: $E[\sigma_t^2]$

$$
\begin{aligned}
E[\sigma_t^2] &= E[\omega + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2] \\
&= \omega + \alpha_1 E[\epsilon_{t-1}^2] + \ldots + \alpha_p E[\epsilon_{t-p}^2] \\
&= \omega + \alpha_1 E[\sigma_{t-1}^2]E[e_{t-1}^2] + \ldots + \alpha_p E[\sigma_{t-p}^2]E[e_{t-p}^2] \\
&= \omega + \alpha_1 E[\sigma_{t-1}^2] + \ldots + \alpha_p E[\sigma_{t-p}^2]
\end{aligned}
$$

If the unconditonal variance exists, $\bar{\sigma}^2 = E[\sigma_{t-j}^2]$ for all $j$, so

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha_1 - \ldots - \alpha_p}$$

## Stationarity

An ARCH(p) proces is stationary if

$$1 - \alpha_1 - \ldots - \alpha_p > 0$$

We also require $\alpha_j > 0$ for all $j$. Why?

Some more analytics: Consider an ARCH(1)

$$
\begin{aligned}
y_t &= \epsilon_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\
\epsilon_t &= \sigma_t e_t \quad e_t \sim N(0,1).
\end{aligned}
$$

Then

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\
\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \sigma_t^2(e_t^2 - 1) \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \nu_t
\end{aligned}
$$

# ARCH(1)

$\nu_t = \sigma_t^2(e_t^2 - 1)$ is the volatility surprise:

1. $E[\nu_t] = 0$
2. $E[\nu_t \nu_{t-j}] = 0$

it's white noise!

This means that $\epsilon_t$ follows an AR(1) like process

$$\epsilon_t^2 - \bar{\sigma}^2 = \sum_{j=0}^{\infty} \alpha_1^j \nu_{t-j}$$

What does the autocorrelation function look like?

# Kurtosis

Another property of ARCH models is that the kurtosis of shocks $\epsilon_t$ is strictly greater than the kurtosis of a normal.

This might seem strange because $\epsilon_t = \sigma_t e_t$ is normal by assumption.

$$\kappa = \frac{E[\epsilon_t^4]}{E[\epsilon_t^2]^2} = \frac{E[\sigma_t^4 E_{t-1}[e_t^4]]}{E[\sigma_t^2 E_{t-1}[e_t^2]]^2} = 3\frac{E[\sigma_t^4]}{E[\sigma_t^2]^2}$$

But we know that $V[\epsilon_t^2] = E[\epsilon_t^4] - E[\epsilon_t^2] = E[\sigma_t^4] - E[\sigma_t^2] \geq 0$.
This means that

$$\frac{E[\sigma_t^4]}{E[\sigma_t^2]^2} \geq 1$$

for $\alpha \neq 0$ we can show this hold strictly.

# GARCH

ARCH models typically require many lags to adequately model conditional variance.

Enter Generalized ARCH (GARCH), introduced by cite:Bollerslev$_{1986}$, a parsimonious way of measuring conditional volatility,

$$
\begin{align}
y_t &= \mu + \epsilon_t \tag{4} \\
\sigma_t^2 &= \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5} \\
\epsilon_t &= \sigma_t e_t \quad e_t \sim N(0,1). \tag{6}
\end{align}
$$

GARCH is an ARCH model with $q$ additional lags of the conditional variance.

Call this a GARCH(p,q) model.

# GARCH(1,1)

$$\begin{aligned} y_t &= \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \quad e_t \sim N(0,1). \end{aligned}$$

We have

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1(\omega + \alpha_1 \epsilon_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\ &= \sum_{j=0}^{\infty} \beta_1^j \omega + \sum_{j=0}^{\infty} \beta_1^j \alpha_1 \epsilon_{t-j-1}^2 \end{aligned}$$

Conditional variance is a constant plus a weighted average of past squared innovations.

Would need many lags to match this with an ARCH.

Parameter restrictions to ensure variances are uniformly positive? Gets very difficult for general GARCH(p,q), see cite:Nelson[1992].

# GARCH(1,1)

Time series model for $\epsilon_t$

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \epsilon_{t-1}^2 + \beta_1 \nu_{t-1} + \nu_t \\
\epsilon_t^2 &= \omega + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 + \beta_1 \nu_{t-1} + \nu_t
\end{aligned}
$$

Instead of AR(1), GARCH(1,1) is transformed into an ARMA(1,1) where $\nu_t = \epsilon_t^2 - \sigma_t^2$ is the volatility surprise.

Unconditional Variance

$$
\bar{\sigma}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}
$$

Autocovariances? Use formulas for autocovariances of ARMA!

# More Flavors of ARCH

Exponential GARCH: [Nelson (1991)] model the *natural log of variance* rather the variance:

$$\ln(\sigma^2) = \omega + \sum_{j=1} \alpha_j \left( \left| \frac{\epsilon_{t-j}}{\sigma_{t-j}} \right| - \sqrt{\frac{2}{\pi}} \right) \sum_{j=1}^{o} \gamma_j \frac{\epsilon_{t-j}}{\sigma_{t-j}} + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2)$$
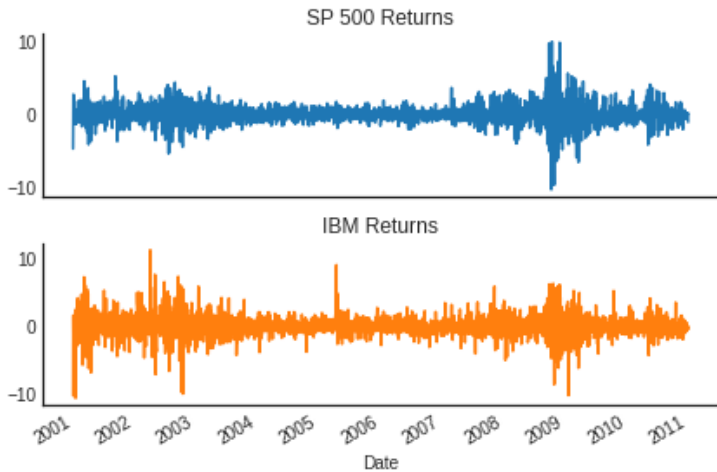
No parameter restrictions!

But, more complicated form
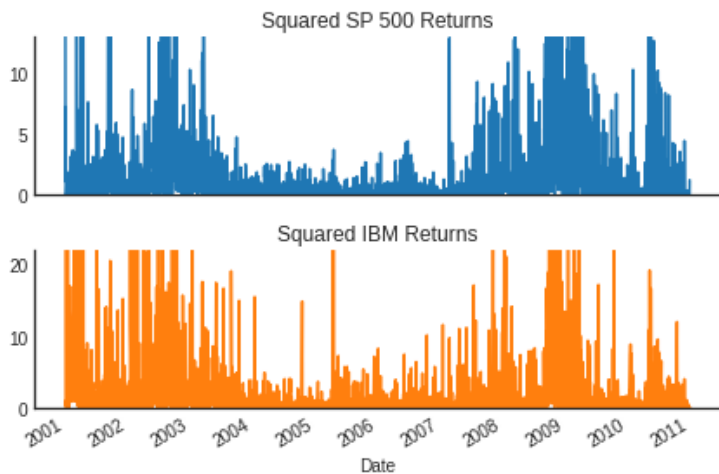
But, possible assymetry.

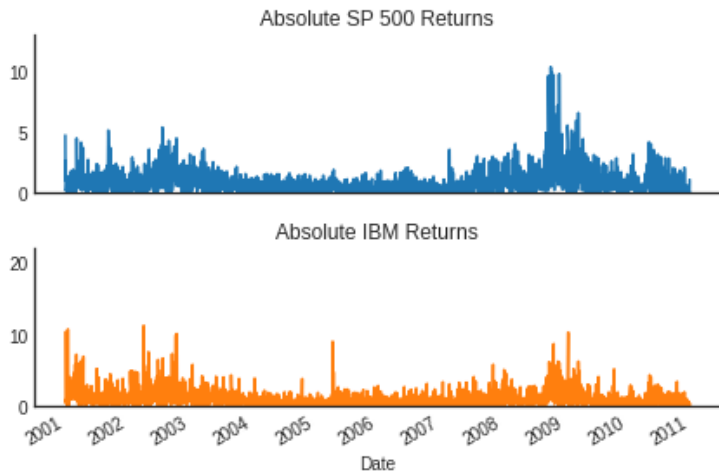*Many other variants possible!*

# Some Data

Daily Data: 2001-2011



SP 500 Returns

IBM Returns

# Evidence of ARCH

# (Less Noisy) Evidence of ARCH



Absolute SP 500 Returns

Absolute IBM Returns

## ARCH(5) for S&P 500

```
----------------------------------------------------------------
NameError                                 Traceback
<ipython-input-33-b18de4cb25c7> in <module>
      1 from      IPython
----> 2 Latex(
NameError: name 'fitted_arch' is not defined
```

# GARCH(1,1) for S&P 500

```
---------------------------------------------------------------------
NameError                                 Traceback
<ipython-input-35-fc0091775d72> in <module
----> 1 Latex(
NameError: name 'fitted_garch' is not defined
```

# EGARCH(1,1,1) for S&P 500

```
^[[0;31m---------------------------------------------------------
^[[0;31mNameError^^[[0m                                   Traceback
^[[0;32m<ipython-input-37-a50630fae910>^^[[0m in ^^[[0;36m<module
^[[0;32m----> 1^^[[0;31m ^^[[0mLatex^^[[0m^^[[0;34m(^^[[0m^^[[0;
^[[0m
^[[0;31mNameError^^[[0m: name 'fitted_egarch' is not defined
```

# ARCH(5) results

```
---------------------------------------------------
NameError                                 Tra
<ipython-input-38-692fe3d6f65d> in <
      1 fig,  ax
----> 2 fitted_arch.

NameError: name 'fitted_arch' is not defined
```

# GARCH(1,1) results

```
^^[[0;31m--------------------------------------------------------
^^[[0;31mNameError^^[[0m                                    Tr
^^[[0;32m<ipython-input-39-d43ebe8803cc>^^[[0m in ^^[[0;36m<
^^[[0;32m----> 1^^[[0;31m ^^[[0mfitted_garch^^[[0m^^[[0;34m.
^^[[0m
^^[[0;31mNameError^^[[0m: name 'fitted_garch' is not defined
```

# EGARCH(1,1,1) results

```
^^[[0;31m---------------------------------------------------------
^^[[0;31mNameError^^[[0m                                Tra
^^[[0;32m<ipython-input-40-310f1d053f53>^^[[0m in ^^[[0;36m<
^^[[0;32m----> 1^^[[0;31m ^^[[0mfitted_egarch^^[[0m^^[[0;34m
^^[[0m
^^[[0;31mNameError^^[[0m: name 'fitted_egarch' is not define
```

# Stochastic Volatility

Consider the stochastic volatlity model

$$y_t = \mu + \sigma_t \epsilon_t \tag{7}$$
$$\ln(\sigma_t^2) = \omega + \alpha_1 \ln(\sigma_{t-1}^2) + \xi_t \tag{8}$$
$$\epsilon_t \sim N(0,1) \text{ and } \xi_t \sim N(0, \sigma_\xi^2) \tag{9}$$

This is different from (G)ARCH: there is *volatility specific* shock.

1. More flexible + link up better to some continuous time asset pricing models.
2. Much harder to estimate
   2.1 GMM: cite:Wiggins[1987].
   2.2 Quasi MLE: cite:Harvey[1994].
   2.3 Bayesian Approach: cite:JPR94.

Connections to GARCH: see cite:Fleming[2003].

# SV vs. GARCH

1. Some studies have found superiority of SV models of GARCH–cite:kimshephardchib1998.
2. But, of course, GARCH models are much easier to estimate.
3. Both are used.

Let's estimate an SV model for the SP 500 returns.

I'm going to do an Bayesian analysis with $\alpha_1 = 1$ (random walk!)
Priors: $\mu \sim Exp(1), \sigma \sim Exp(1/0.02)$,

I'm going to use an MCMC sampler designed for this kind of model.

It takes about 20 minutes to estimate [GARCH is basically instant]

## SV vs. EGARCH

```
^^[[0;31m-------------------------------------------------------
^^[[0;31mNameError^^[[0m                                      Tra
^^[[0;32m<ipython-input-42-b5076a34bed0>^^[[0m in ^^[[0;36m<
^^[[0;32m----> 1^^[[0;31m ^^[[0msv^^[[0m ^^[[0;34m=^^[[0m ^
^^[[0m^^[[1;32m      2^^[[0m ^^[[0mmax^^[[0m ^^[[0;34m=^^[[0m
^^[[1;32m      3^^[[0m ^^[[0mfitted_egarch^^[[0m^^[[0;34m.^^
^^[[1;32m      4^^[[0m ^^[[0mmax^^[[0m^^[[0;34m.^^[[0m^^[[0m

^^[[0;31mNameError^^[[0m: name 'trace' is not defined
```

# SV vs. EGARCH (2005)

```
^^[[0;31m---------------------------------------------------
^^[[0;31mNameError^^[[0m                                    Tra
^^[[0;32m<ipython-input-43-fe4883815d9a>^^[[0m in ^^[[0;36m<
^^[[0;32m----> 1^^[[0;31m ^^[[0msv^^[[0m ^^[[0;34m=^^[[0m ^^
^^[[0m^^[[1;32m      2^^[[0m ^^[[0max^^[[0m ^^[[0;34m=^^[[0m
^^[[1;32m      3^^[[0m ^^[[0mfitted_egarch^^[[0m^^[[0;34m.^^
^^[[1;32m      4^^[[0m ^^[[0max^^[[0m^^[[0;34m.^^[[0m^^[[0m

^^[[0;31mNameError^^[[0m: name 'trace' is not defined
```

# A famous SV paper...

cite:STOCK$_{2007}$: "Why Has Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking*.

Has inflation become harder to forecast recently?
1. No, MSE of forecasts has gone down.
2. Yes, improvement of "structural" models relative to univariate ones have diminished.

Propose an Unobserved Components-Stochastic Volatility (UC-SV) model for inflation:

$$
\begin{aligned}
\pi_t &= \tau_t + \eta_t, \quad \eta_t \sim N(0, \sigma_{\eta,t}^2) \\
\tau_t &= \tau_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{\epsilon,t}^2) \\
\ln \sigma_{\eta,t}^2 &= \ln \sigma_{\eta,t-1}^2 + \nu_{\eta,t}, \nu_t \sim N(0, \gamma^2) \\
\ln \sigma_{\epsilon,t}^2 &= \ln \sigma_{\epsilon,t-1}^2 + \nu_{\epsilon,t}, \sim N(0, \gamma^2)
\end{aligned}
$$

$\tau_t$ = permanent (stochastic) trend inflation
$\eta_t$ = transitory component.

# Inflation



Inflation

## Estimation

There are no parameters to estimate, as SW set $\gamma = 0.2$.

Still have to filter the volatilities $\{\sigma_{\eta,t}\}_{t=1}^T$ and $\{\sigma_{\epsilon,t}\}_{t=1}^T$.

SW use a Markov-chain Monte Carlo (MCMC) technique to do this.

Compare this to a time-varying Integrated Moving Average (IMA) model:

$$\Delta\pi_t = (1 - \theta_t L)e_t, \quad e_t \sim iid(0, \sigma_e^2). \tag{10}$$

UC-SV model implies that:

$$\theta_t = \frac{1 - \sqrt{1 - 4\alpha_t^2}}{2\alpha}, \quad \alpha_t = \frac{\sigma_{\eta,t}^2}{\sigma_{\eta,t}^2 + \sigma_{\epsilon,t}^2}$$

## Results

```
^^[[0;31m---------------------------------------------------------
^^[[0;31mNameError^^[[0m                                       Tra
^^[[0;32m<ipython-input-45-6447f4614425>^^[[0m in ^^[[0;36m<
^^[[0;32m----> 1^^[[0;31m ^^[[0msigma_eps^^[[0m ^^[[0;34m=^
^^[[0m^^[[1;32m      2^^[[0m ^^[[0msigma_eta^^[[0m ^^[[0;34m
^^[[1;32m      3^^[[0m ^^[[0mtheta^^[[0m ^^[[0;34m=^^[[0m ^
^^[[1;32m      4^^[[0m ^^[[0mtitle^^[[0m ^^[[0;34m=^^[[0m ^
^^[[1;32m      5^^[[0m ^^[[0mfig^^[[0m^^[[0;34m,^^[[0m^^[[0m

^^[[0;31mNameError^^[[0m: name 'tracesw' is not defined
```

# Discussion

- SD of permanent component: 1970s through 1983 = high volatility, 1950s-1960s = moderate volatility, post-83 = low volatility

- SD of transitory component: little change

- IMA estimate: moderate in early sample, falling in the 70s, increasing sharply thereafter.

# Markov Switching
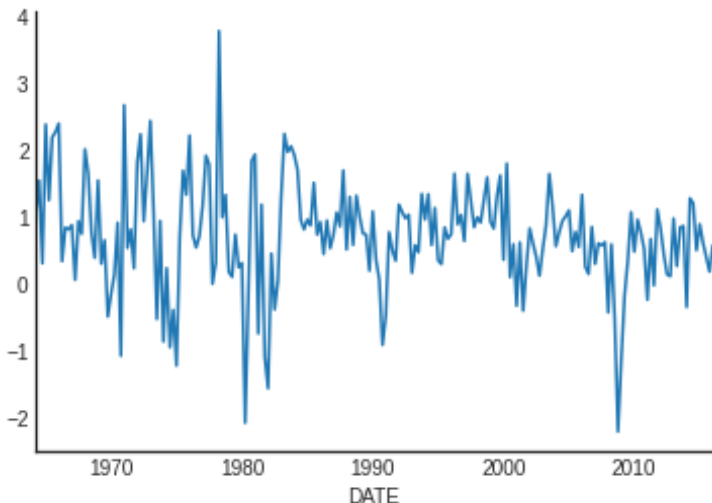
Markov Switching – *exogenous* switching in parameters

Useful references:
- ▶ Hamilton, Chapter 22.
- ▶ cite:KimNelson1999

# US GDP growth

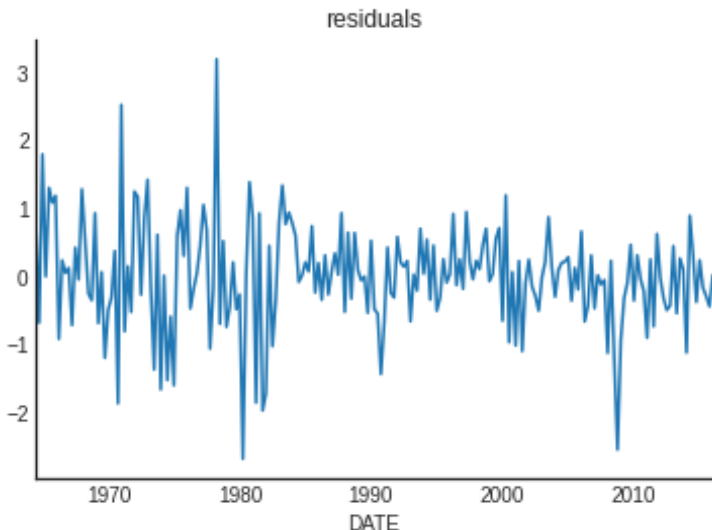Suppose we want to model U.S. GDP growth as a stationary AR(1):

$$y_t = \mu + \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2).$$

# Estimation Results

We get point estimates: $(1 - \hat{\rho})\hat{\mu} = 0.5$, $\hat{\rho} = 0.31$. How does it fit? Look at residuals.



residuals

# Residuals, pre and post 1984

1. The standard deviation is considerable smaller post 1984 (even with GR!)
2. While this is only suggestive, indicates there was a *break* in series.
3. What has broken once, can break again. How to incorporate?

# Markov Switching

Let's consider the following representation:

$$y_t = \mu(s_t) + \rho y_{t-1} + \epsilon_t, \quad \epsilon_t N(0, \sigma^2(s_t))$$

The mean and variance are now functions of an unobserved discrete random variable: $s_t$.

Call realization of $s_t$ the state (or regime) that discrete random variable takes.

- $s_t = 1$, $\mu(s_t) = \mu_1, \sigma^2(s_t) = \sigma_1^2$
- $s_t = 2$, $\mu(s_t) = \mu_2, \sigma^2(s_t) = \sigma_2^2$

We need an description for time series. A simple model is a Markov chain!

Let's also set $\rho = 0$ for now, for simplicity.

# Markov Chains, continued

We talked about Markov Chains when we talked about Markov Chain Monte Carlo. Let's review.

Suppose $s_t$ is an RV that takes values in $\{1, 2, \ldots, N\}$. The probabilitiy distribution for $s_t$ depends only one the past through it's most recent realization $s_{t-1}$.

$$P(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots) = P(s_t = j | s_{t-1} = i) = p_{ij}$$

$p_{ij}$ is the probability of state $j$given state $i$ last period.

Note that

$$\sum_{j=1}^{N} p_{ij} = 1, \text{ for } i = 1, \ldots, N.$$

## Markov Chains, continued

Stack these probablity in a matrix.

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{N1} & p_{N1} & \cdots & p_{NN} \end{bmatrix} \tag{11}$$

Let $\xi_t = [1, 0, 0, \ldots, 0]'$ when $s_t = 1$, $\xi_t = [0, 1, 0, \ldots, 0]'$ when $s_t = 2$, and so on. This means that

$$E[\xi_{t+1}|\xi_t] = E[\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots] = P\xi_t.$$

We represent our Markov Chain as

$$\xi_{t+1} = P\xi_t + \nu_t, \quad \nu_t = \xi_{t+1} - E[\xi_{t+1}|\xi_t]$$

$\nu_t$ is an Martingale Difference Series: mean zero and impossible to forecast using previous states.

# Forecasting with Markov Chain

$$\xi_{t+m} = \nu_{t+m} + P\nu_{t+m-1} + P^2\nu_{t+m-2} + \ldots + P^m\xi_t$$

This means that

$$E[\xi_{t+m}|\xi_t] = P^m\xi_t.$$

The probability that a state from regime $i$ will be followed $m$ periods later by a realization of state $j$ is given by the $j$th row, $i$th column element of $P^m$.

Concepts:

1. *reducibility:* if the chain is vanishing, in the sense that once you visit a state, you will not return to some other states. Not reducible? {*irreducible*

2. *ergodicity:* An irreducible MC is ergodoc, is ergodic is there exists a probability vector $\pi$ such that $P\pi = \pi$.

# More MC

1. An ergodic Markov Chain is a covariance stationary process,
2. But, the VAR has a unit root in it!
3. Magic: the variance matrix of $\nu_t$ is singular.

Calculating ergodic probabilty:

$$P\pi = \pi \text{ and } 1'\pi = 1.$$

So

$$\begin{bmatrix} I - P \\ 1' \end{bmatrix} \pi = A\pi = e_{N+1}$$

This means that $\pi = (A'A)^{-1}A'e_{n+1}$.

Periodic markov chain: there is more than one $\pi$ s.t. $P\pi = \pi$.

# Analyzing Mixtures of Normals

With $\rho = 0$, $y_t$ is normally distributed conditional on $s_t$.

$$p(y_t | s_t = j; \theta) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}}$$

We also know that $P(A \text{ and } B) = P(A|B)p(B)$. So:

$$p(y_t, s_t = j; \theta) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}} \times p(s_t = j)$$
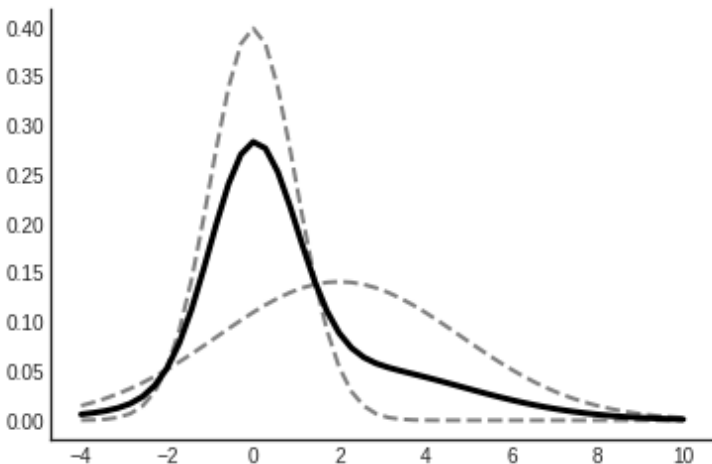
If $p(s_t = j; \theta) = \pi_j$, we have

$$p(y_t; \theta) = \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_j} \exp{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}} \times \pi_j, \quad p(Y_{1:T}; \theta) = \prod_{t=1}^{T} p(y_t; \theta)$$

This is a mixture of normals.
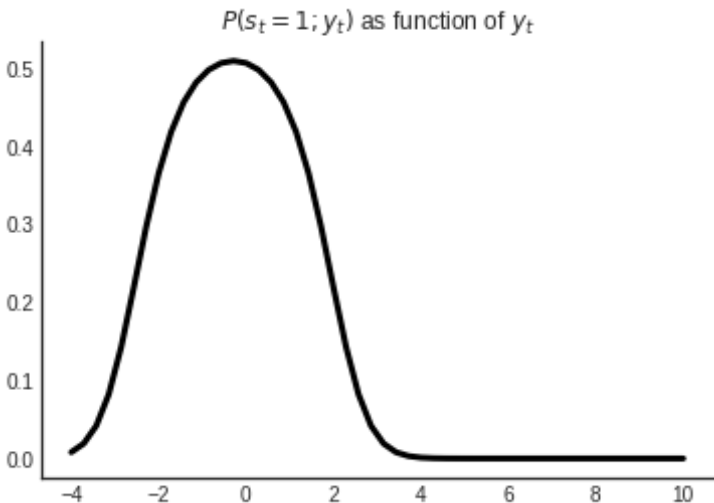
# Mixture of $N(0,1)$ and $N(2,8)$, $\pi_1 = 0.6$.



you can make some pretty crazy (any) distributions with mixture of normals.

# More on Mixtures

How to infer which state I'm in? *Bayes Rule!*

$$p(s_t = j | y_t; \theta) = \frac{p(y_t | s_t = j; \theta) p(s_t = j; \theta)}{p(y_t; \theta)}$$

$P(s_t = 1; y_t)$ as function of $y_t$

## How to estimate Mixtures

One can show that the MLE's of $\theta = [\mu, \sigma, \pi]$ are

$$
\begin{aligned}
\hat{\mu}_j &= \frac{\sum_{t=1}^{T} y_t p(s_t = j | y_t, \hat{\theta})}{\sum_{t=1}^{T} p(s_t = j | y_t, \hat{\theta})}, \quad j = 1, \ldots, N \\
\hat{\sigma}_j^2 &= \frac{\sum_{t=1}^{T} (y_t - \hat{\mu}_j)^2 p(s_t = j | y_t, \hat{\theta})}{\sum_{t=1}^{T} p(s_t = j | y_t, \hat{\theta})}, \quad j = 1, \ldots, N \\
\hat{\pi} &= \frac{1}{T} \sum_{t=1}^{T} p(s_t = j | y_t, \hat{\theta})
\end{aligned}
$$

Recall, we get $p(s_t = j | y_t, \theta)$ from applying Bayes rule.

This forms a system of nonlinear equations, which in principal can be solved.

Another way...

# The EM Algorithm

A simple iterative procedure:

1. Guess arbitrary $\hat{\theta}^0$,
2. Plugin in to RHS of above equations, get $\hat{\theta}^1$
3. if $|\hat{\theta}^1 - \hat{\theta}^0| < \epsilon$, you've found a maximum
4. otherwise continuing iterating

This is called the *EM Algorithm.*

1. Each iteration increases the value of the likelihood function
2. Usually works very fast
3. Not guaranteed to find a global max, but generally robust.

# Issues in optimization

1. if $\hat{\mu}_j = y_t$ for some j and $y_t$, we're in trouble, because the max will send $\sigma_j \to 0$.
2. This makes the log-likelihood go to infinity
3. More generally: there is a labeling issue: $(\mu_1, \sigma_1^2)$ could easily be called $(\mu_2, \sigma_2^2)$ and vice versa.
4. This calls for augmenting the problem with more information.
   4.1 Coherent way: Bayesian Approach, see cite:KimNelson1999.
   4.2 Ad hoc way: augment the likliehood with a penalty function to enforce priors views and make pathologies less likely. See cite:Hamilton.

## Markov Switching

Markov Switching models are more complicated versions of the mixture model.

Instead of just $\pi = [\pi_1, \ldots, \pi_N]$ we need to estimate $P$

$$
P = \begin{bmatrix}
p_{11} & p_{21} & \cdots & p_{N1} \\
p_{12} & p_{22} & \cdots & p_{N2} \\
\vdots & \vdots & \cdots & \vdots \\
p_{N1} & p_{N1} & \cdots & p_{NN}
\end{bmatrix}
\tag{12}
$$

Key item is again: distribution of states conditional on data.

Now we need to consider entire history of data, because of Markovian structure.

$$
\hat{\xi}_{t|t} = [p(s_t = 1|Y_{1:t}; \theta), \ldots, p(s_t = N|Y_{1:t-1}; \theta)]'
$$

We can also think about the forecast of this probability distribution

$$
\hat{\xi}_{t|t-1}.
$$

## Deriving the likelihood

The joint distribution of $y_t$ and $s_t$ given $Y_{1:t-1}$ is given by

$$p(y_t, s_t = j | Y_{1:t-1}; \theta) = p(y_t | s_t = j, Y_{1:t-1}; \theta) p(s_t = j | Y_{1:t-1}; \theta)$$

Let $\eta_t = [p(y_t | s_t = 1, Y_{1:t-1}; \theta), \ldots, p(y_t | s_t = N, Y_{1:t-1}; \theta)]'$.

Recall $\hat{\xi}_{t|t-1} = [p(s_t = 1 | Y_{1:t-1}; \theta), \ldots, p(s_t = N | Y_{1:t-1}; \theta)]'$

Then $p(y_t, s_t = j | Y_{1:t-1}; \theta) = \hat{\xi}_{t|t-1} \odot \hat{\eta}_t$.

$$p(y_t | Y_{1:t-1}; \theta) = \sum_{j=1}^{N} p(y_t, s_t = j | Y_{1:t-1}; \theta) = 1'(\hat{\xi}_{t|t-1} \odot \eta_t)$$

The log likelihood is

$$\sum_{j=1}^{T} \log p(y_t | Y_{1:t-1}; \theta) = \sum_{t=1}^{T} \log \left( 1'(\hat{\xi}_{t|t-1} \odot \eta_t) \right)$$

Now to get a recursion for $\hat{\xi}_{t|t-1}$

# Recursion for $\hat{\xi}_{t|t-1}$

Use Bayes rule

$$p(s_t = j | Y_{1:T}; \theta) = \frac{p(y_t | s_t = j; \theta) p(s_t = j | Y_{1:t-1}; \theta)}{p(y_t | Y_{1:t-1}; \theta)}$$

Which is just

$$p(s_t = j | Y_{1:T}; \theta) = \hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{1'(\hat{\xi}_{t|t-1} \odot \eta_t)}$$

To get forecast:

$$E_t[\xi_{t+1} | Y_{1:T}] = P E_t[\xi_t | Y_{1:T}] + E_t[\nu_t | Y_{1:T}] = P \hat{\xi}_{t|t}$$

And we're done!

# Smoothed Probabilities

Kim Algorithm:

Do this stuff "backwards":

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot [P'[\hat{\xi}_{t+1}/\hat{\xi}_{t+1|t}]]$$

Estimation

1. Hamilton: If initial probablities are unrelated

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} p(s_t = j, s_t = i | Y_{1:T}; \theta)}{\sum_{t=2}^{T} p(s_{t-1} = i; Y_{1:T}\theta)}$$

2. Otherwise, a bit more complicated
3. either way use EM algorithm.

## Back to Example

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dep. Variable:** | | GDPC1 | | **No. Observations:** | | 207 |
| **Model:** | | MarkovAutoregression | | **Log Likelihood** | | -220.722 |
| **Date:** | | Thu, 13 Mar 2025 | | **AIC** | | 455.444 |
| **Time:** | | 17:22:53 | | **BIC** | | 478.773 |
| **Sample:** | | 04-01-1964 | | **HQIC** | | 464.878 |
| | | - 01-01-2016 | | | | |
| **Covariance Type:** | | approx | | | | |

| | coef | std err | z | P> |z| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 0.7393 | 0.066 | 11.280 | 0.000 | 0.611 | 0.868 |
| **sigma2** | 0.2301 | 0.035 | 6.557 | 0.000 | 0.161 | 0.299 |
| | coef | std err | z | P> |z| | [0.025 | 0.975] |
| **const** | 0.7114 | 0.168 | 4.239 | 0.000 | 0.382 | 1.040 |
| **sigma2** | 1.1565 | 0.197 | 5.862 | 0.000 | 0.770 | 1.543 |
| | coef | std err | z | P> |z| | [0.025 | 0.975] |
| **ar.L1** | 0.2834 | 0.074 | 3.847 | 0.000 | 0.139 | 0.428 |
| | coef | std err | z | P> |z| | [0.025 | 0.975] |
| **p[0->0]** | 0.9792 | 0.018 | 53.932 | 0.000 | 0.944 | 1.015 |
| **p[1->0]** | 0.0289 | 0.027 | 1.071 | 0.284 | -0.024 | 0.082 |

Warnings:

[1] Covariance matrix calculated using numerical (complex-step) differentiation.

# Smoothed Probabilities

# References I