Hierarchical Mixtures of AR Models for Financial Time Series Analysis

Carmen Vidal[1] & Alberto Suárez [1,2]

(1) Computer Science Dpt., Escuela Politécnica Superior

(2) Risklab Madrid

Universidad Autónoma de Madrid (Spain)

alberto.suarez@ii.uam.es

Financial time series

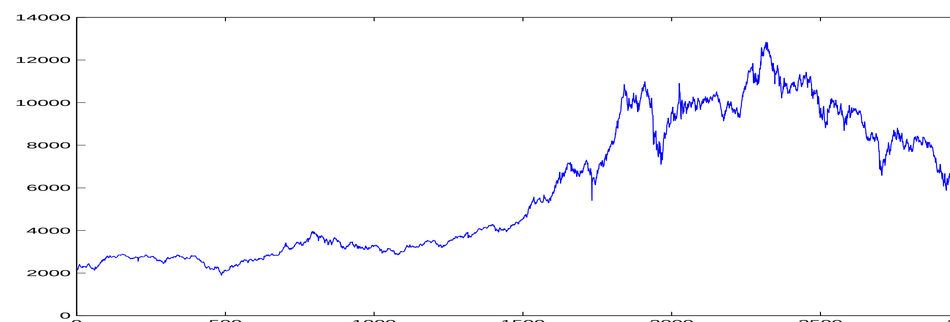
- Time series of assets are highly irregular
  - If market efficiency hypothesis is correct they are also unpredictable.
- Time series of assets are non-stationary

  They are usually transformed in log-returns, or, for short periods of time, in relative returns

- Asset returns exhibit deveations from normality
  - Leptokurtic: Heavy tails
  - Heteroskedastic: Volatility clustering

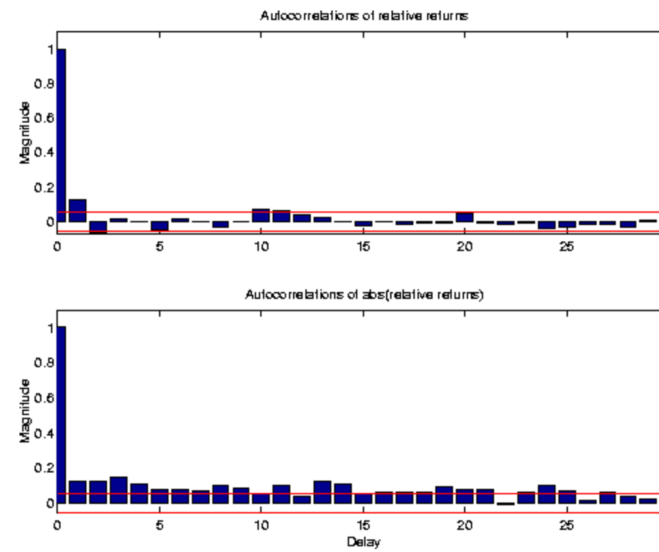
Financial time series modelling/ analysis

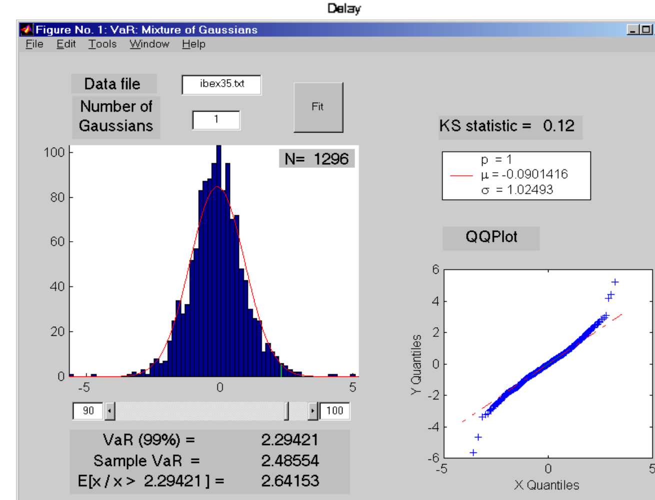
- Modelling finacial time-series is not easy
  - Natural sciences
    - Not reproducible
    - Underlying model?
  - Inductive / statistical learning
    - Small data sets
    - Complex data
      - Non-linear
      - Non-stationarity
      - Non-gaussian
      - Heteroskedastic



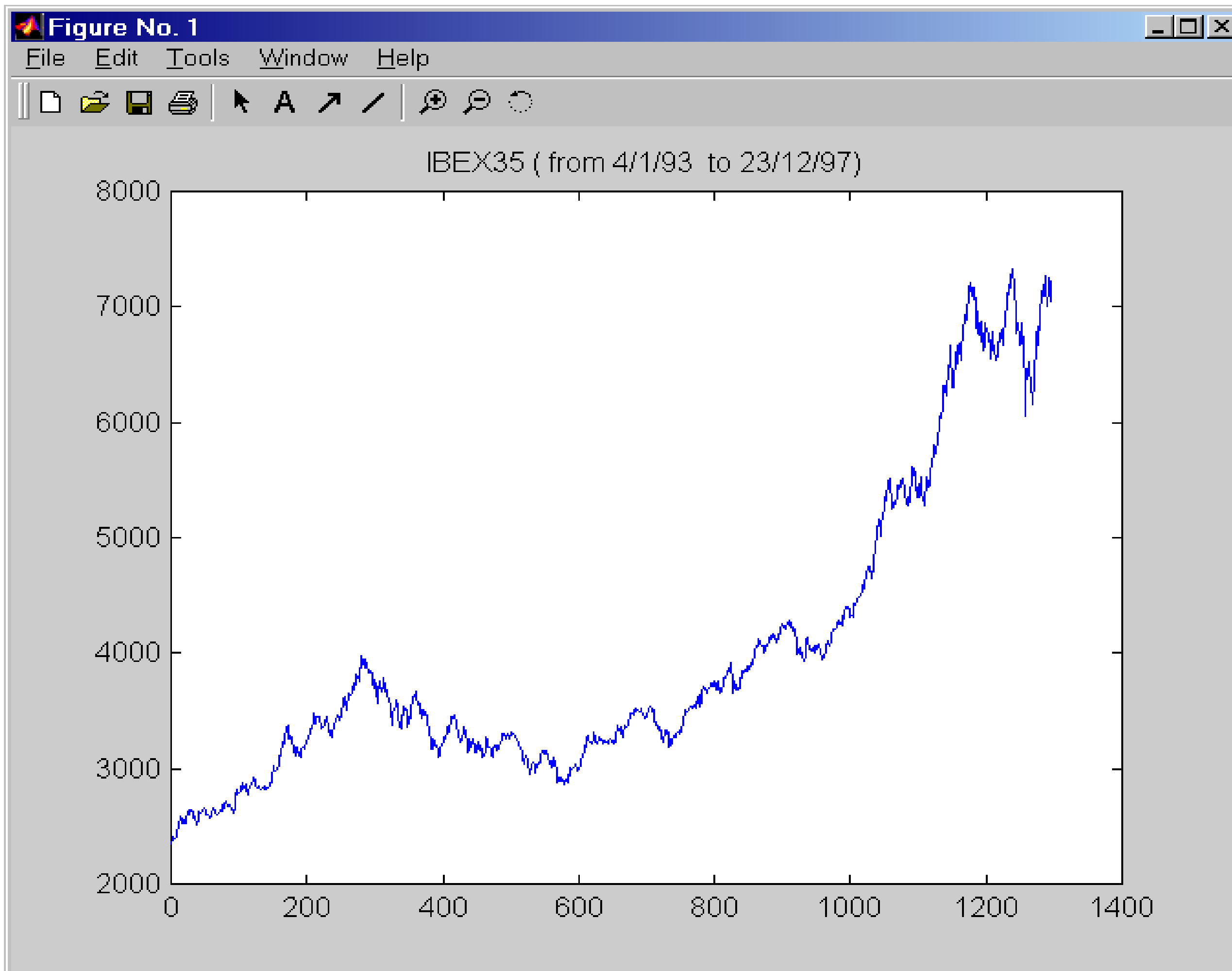
Two stylized facts (Timo Teräsvirta)

- Returns exhibit two empirically observed features:
  - Correlations
    - Short term for the returns
    - Medium term for absolute values of returns
  - Leptokurtosis
    - Heavy tails
    - Extreme events

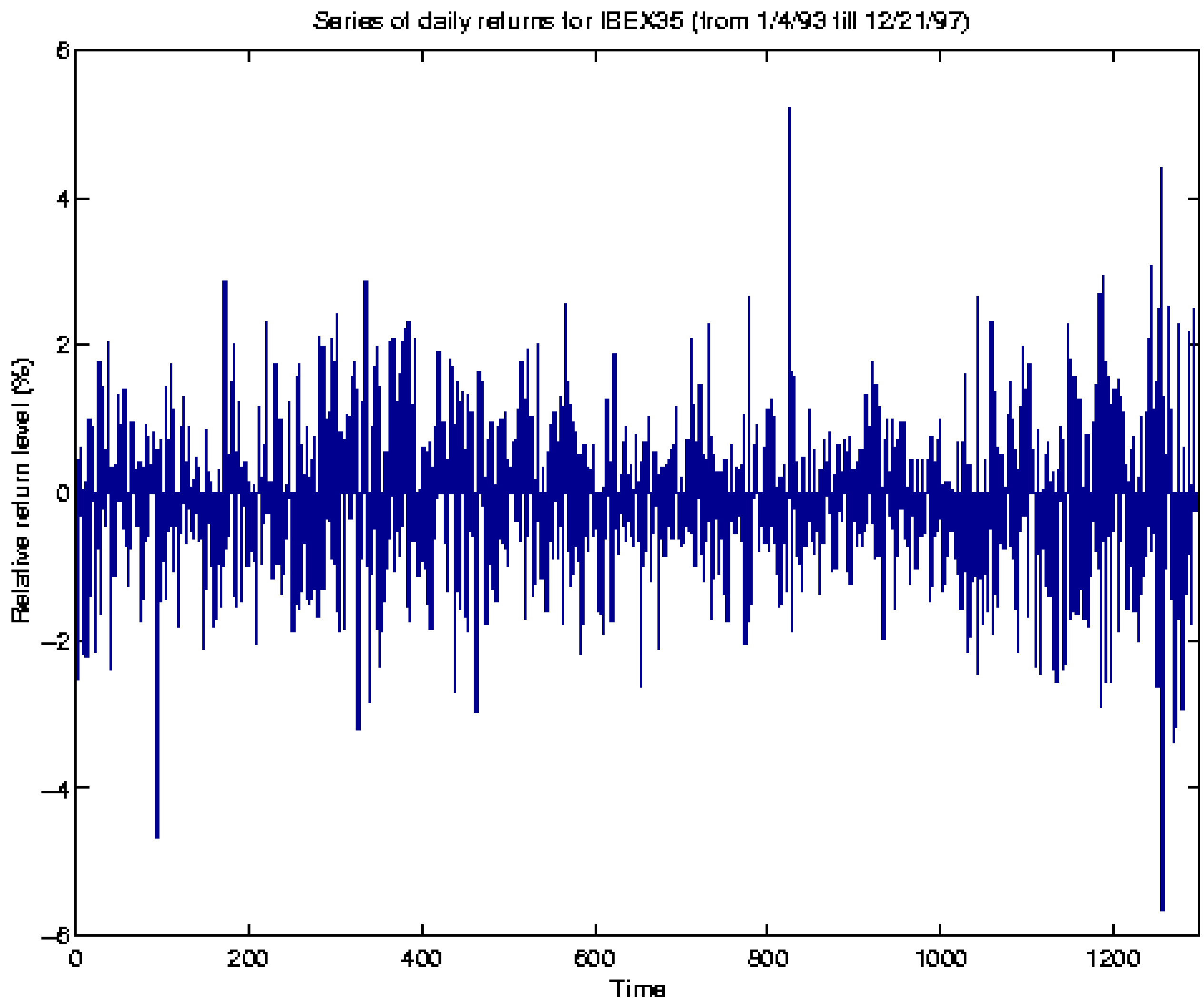




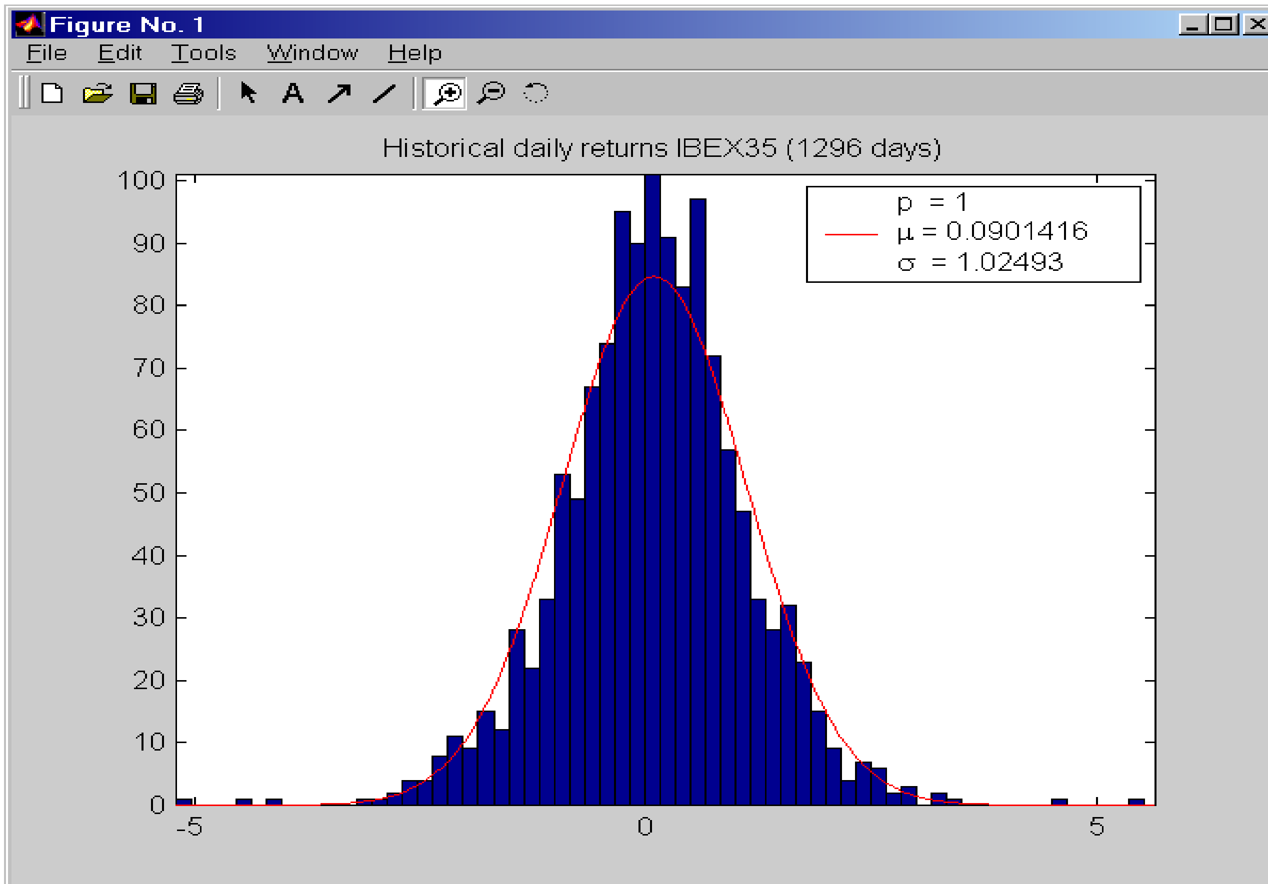
An example: IBEX35



Daily returns: IBEX35 (5 years)



Daily-returns distribution



Black-Scholes theory

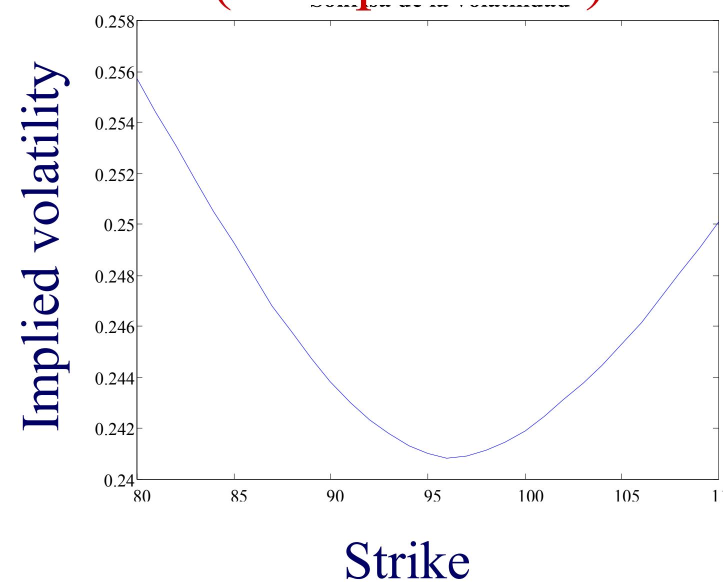
- In theory: Markets are efficient
  - Absence of arbitrage opportunities.
  - No systematic trends.
  - Very short term memory.
- Model: Black-Scholes
  - Log of daily returns of an asset are distributed according to a normal distribution.
  - Two parameters:
    - Risk free **interest rate**.
    - **Volatility [free parameter]**

Is Black-Scholes a good model?

- **Advantages**
  - Simple minimal model with only **one free parameter,** the volatility**.**
  - **Good** pricing **accuracy** for **at-the-money** options.
  - **Analytic** pricing **formulas** for simple derivatives.
- **Drawbacks:**
  - Incorrect pricing formulas for:
    - Deep in-the-money or out-of-the-money
    - Short-term (less than a month) orptions
    - Options on underlying with very low or very high volatility.

This is reflected in the fact that **implied volatility is not constant** [Volatility smile]

Volatility **smile**
(European call)



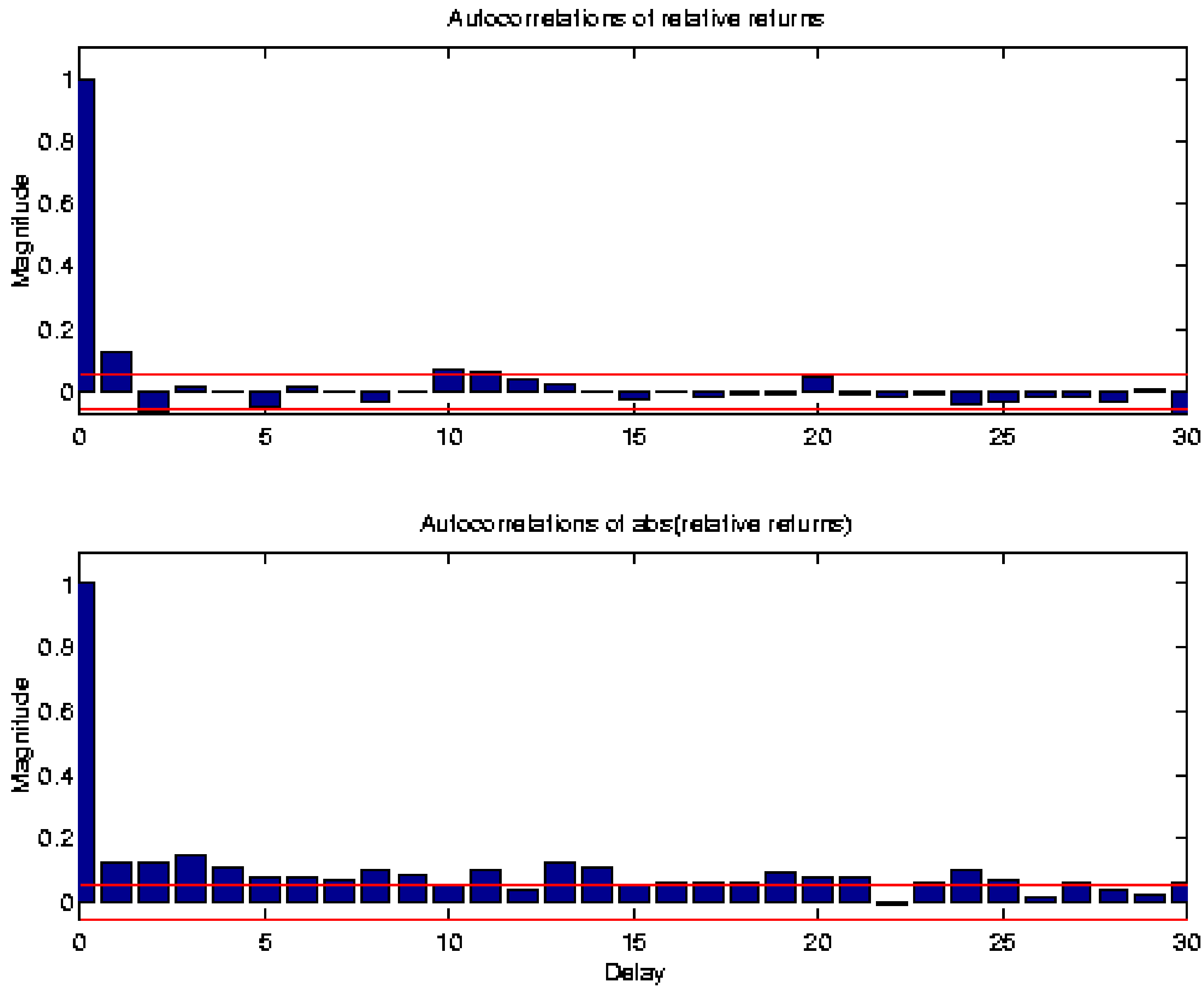
Beyond Black-Scholes

- In practice markets are
  - Not efficient: Memory effects (short/long term?).
  - Very unpredictable (at least sometimes)
    - **Extreme events** are more frequent than what the Black-Scholes models predicts.
    - Occurrence of **crashes**.
    - **Changes** in economic paradigm.
  - Market friction: Transaction costs, lack of liquidity, dividends, etc.

Heteroskedasticity + heavy tails

- Need more sophisticated model
  - Parametric models: Generalizations of Blak-Scholes.
  - Non-parametric models: **Neural networks, Mixture models**

Memory effects (IBEX 35)



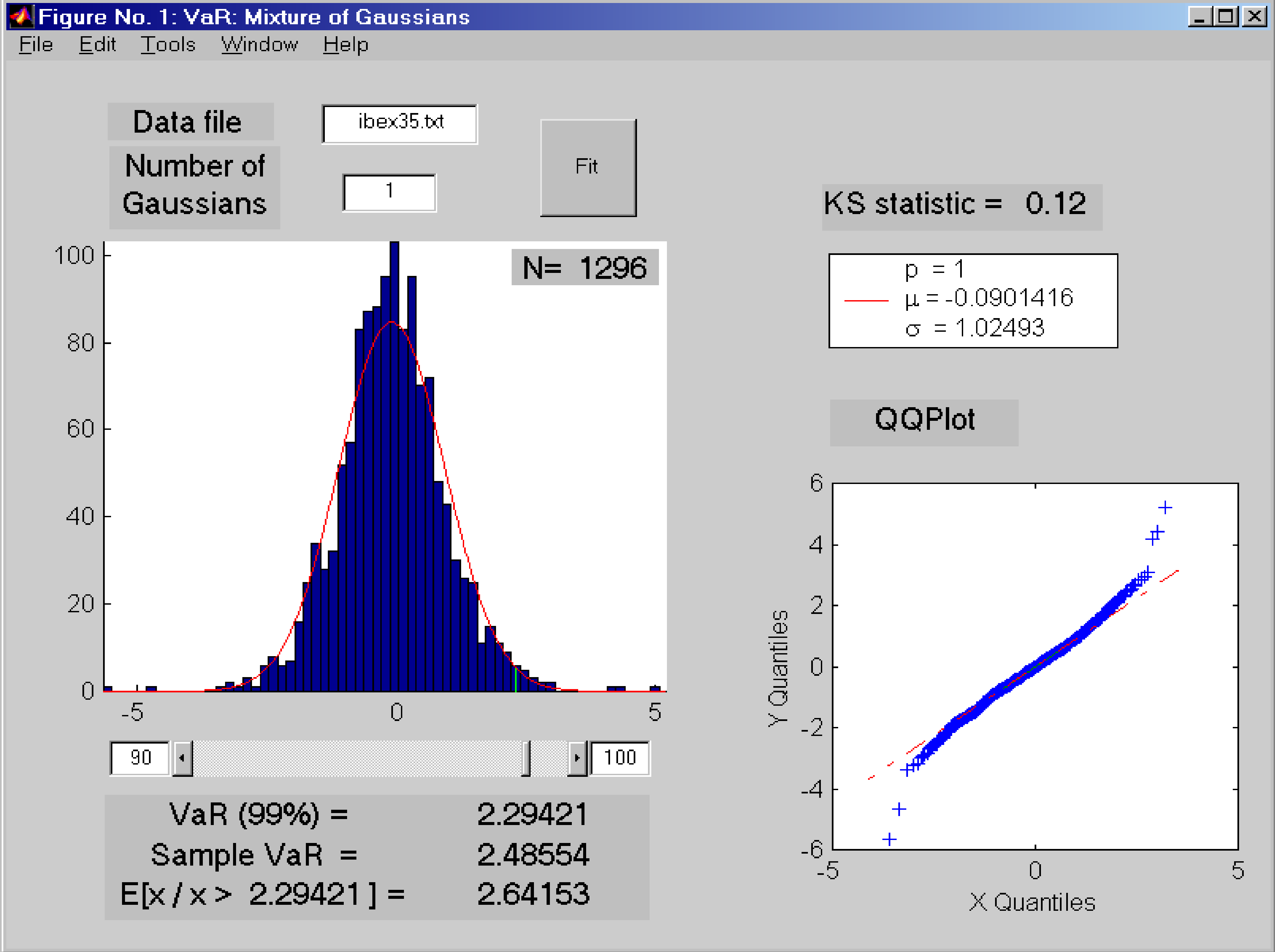
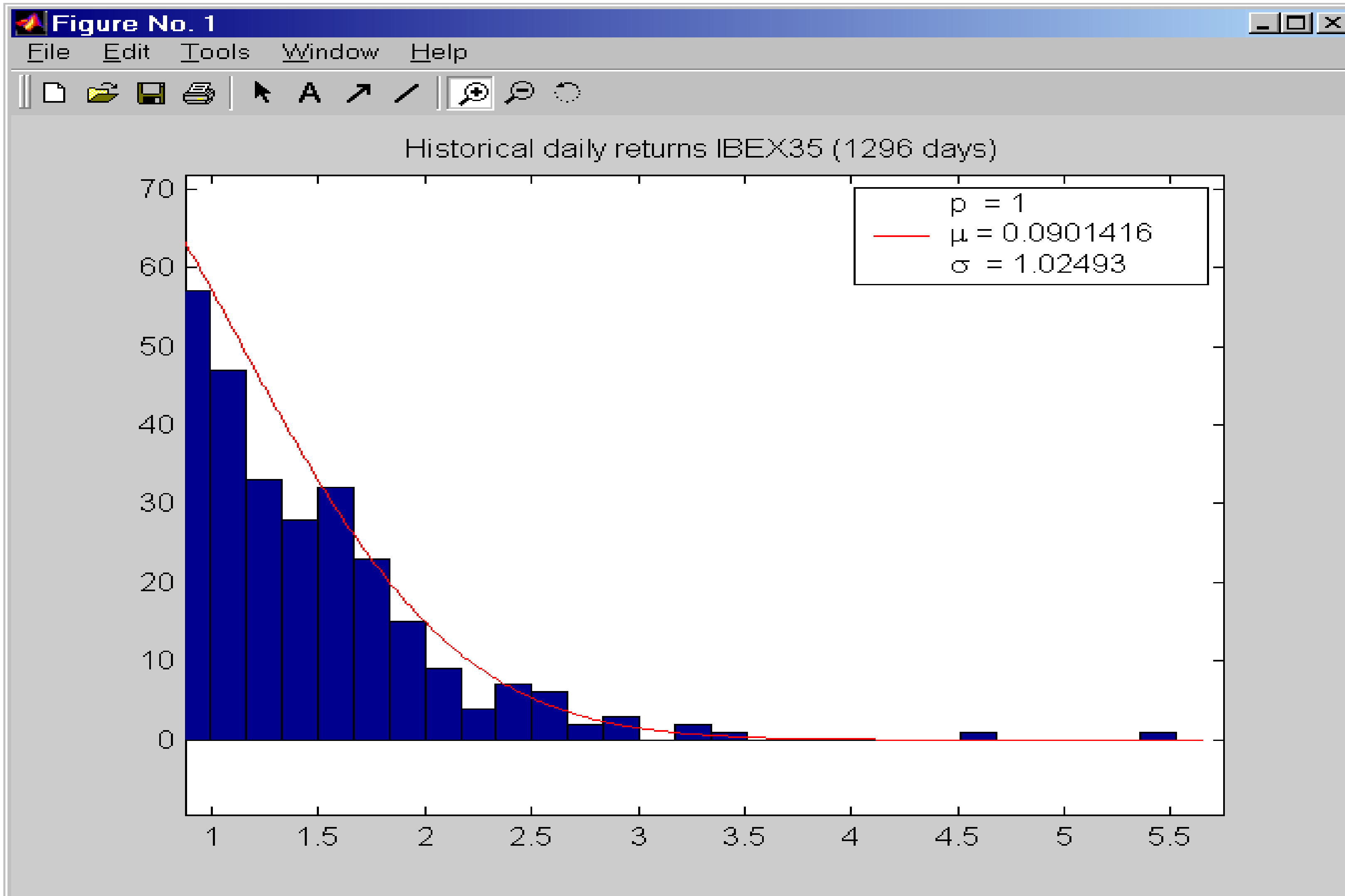


Figure No. 1: VaR: Mixture of Gaussians
File Edit Tools Window Help
Data file
ibex35.txt
Number of Gaussians
1
Fit
KS statistic = 0.12
N= 1296
p = 1
μ = -0.0901416
σ = 1.02493
QQPlot
100
80
60
40
20
0
-5
0
5
90
100
6
4
2
0
-2
-4
-6
Y Quantiles
X Quantiles
VaR (99%) = 2.29421
Sample VaR = 2.48554
E[x / x > 2.29421] = 2.64153

Failure of normal model: Heavy tails



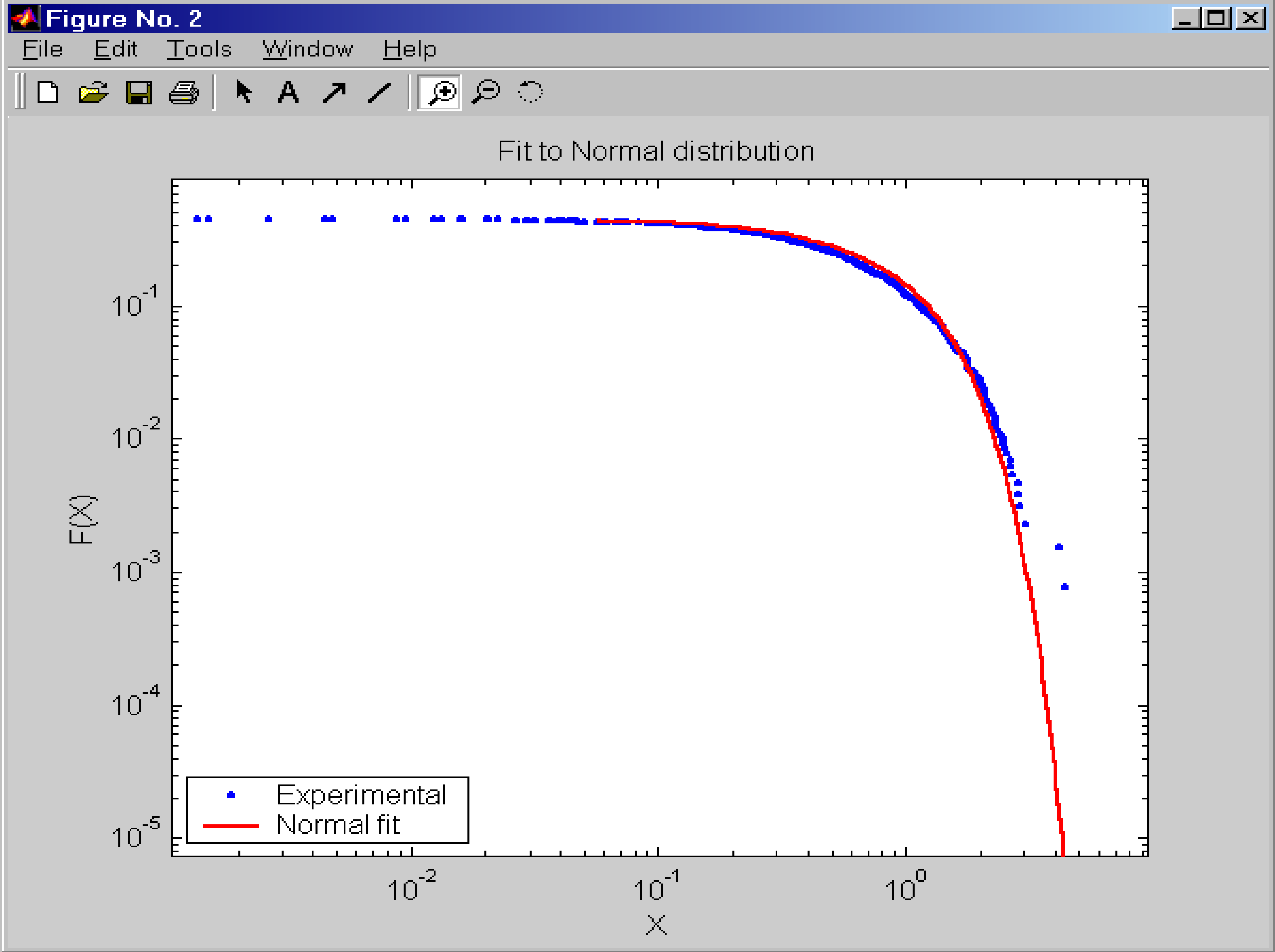
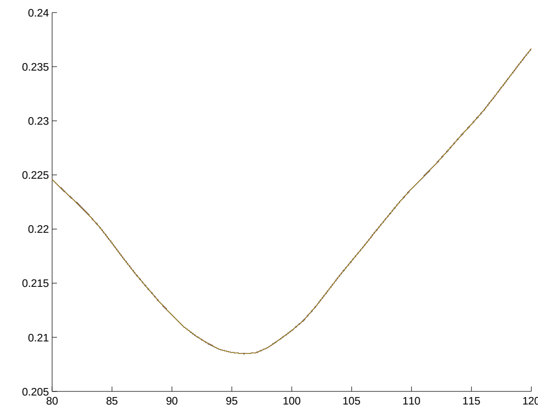


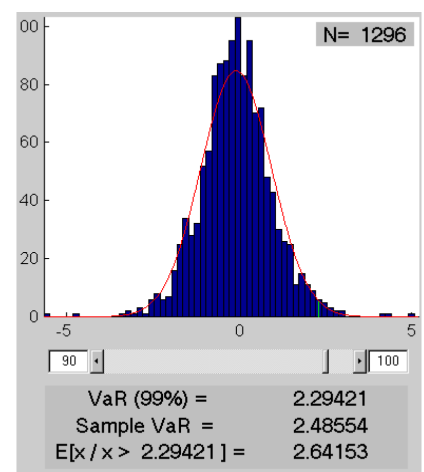
Figure No. 2
File Edit Tools Window Help
Fit to Normal distribution
F(X)
X
Experimental
Normal fit

Empirical evidence for leptokurtosis

- Volatility smiles and smirks
  - Black-Scholes is insufficient to account for time evolution of underlying.



- Incremented risk
  - Multiplicative factor in market Risk estimates (Basel Accord 1988, 1996 ammendment)



Time series analysis

- Consider the time series

$$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t, \ldots, \mathbf{X}_T$$

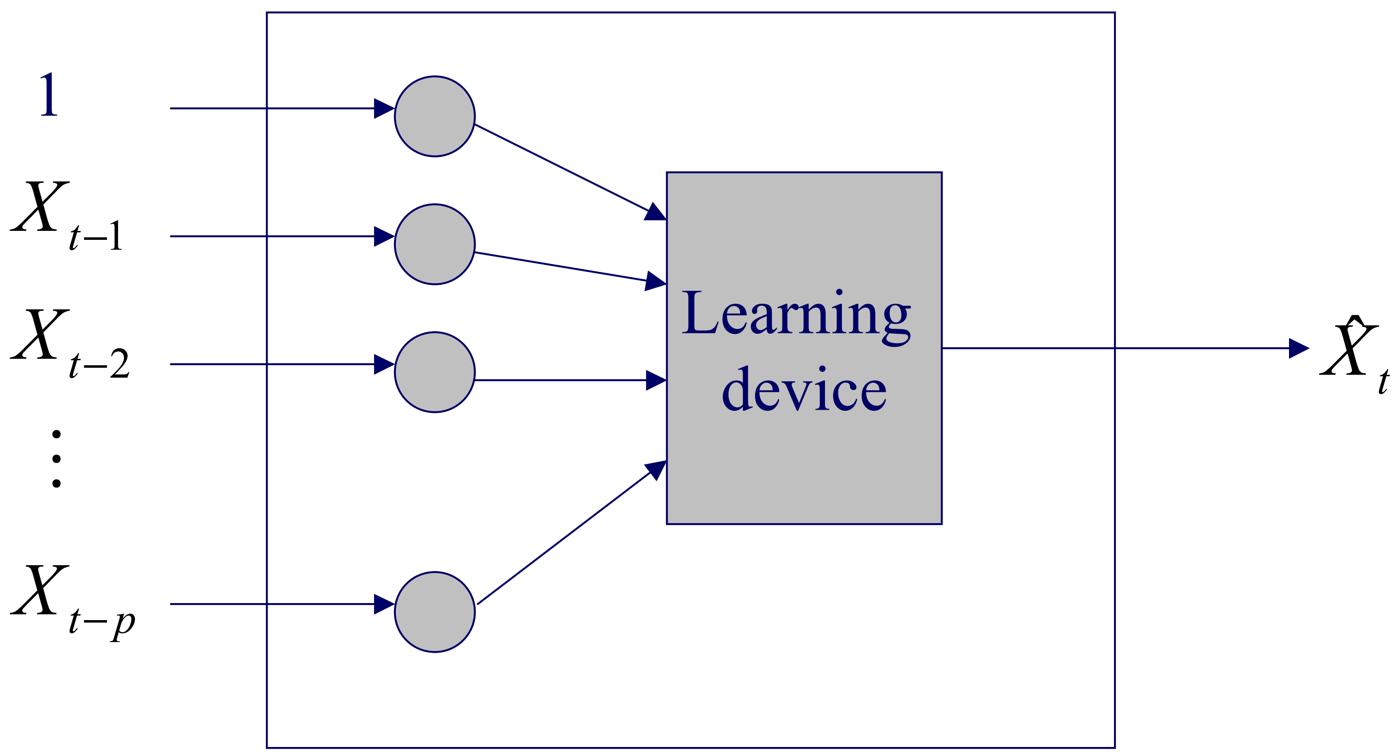
- **Time series analysis**
  - Forecasting $\hat{\mathbf{X}}_{\mathbf{t+d}} = F(\mathbf{X}_\mathbf{t}, \mathbf{X}_{\mathbf{t-1}}, \ldots; \theta_t);$
  - Classification $Class = F(\mathbf{X}_\mathbf{t}, \mathbf{X}_{\mathbf{t-1}}, \ldots; \theta_t)$
  - Modelling $P(\mathbf{X}_{\mathbf{t+d}} \mid \mathbf{X}_\mathbf{t}, \mathbf{X}_{\mathbf{t-1}}, \ldots; \theta_t)$
- These problems are closely related to each other:

$$\mathbf{X}_{\mathbf{t+d}} = F(\mathbf{X}_\mathbf{t}, \mathbf{X}_{\mathbf{t-1}}, \ldots; \theta_t) + \varepsilon_{t+d};$$

$$P(\varepsilon_{\mathbf{t+d}} \mid \mathbf{X}_\mathbf{t}, \mathbf{X}_{\mathbf{t-1}}, \ldots; \theta_t)$$

Time series prediction: a Learning view

⌘Network model for time-series prediction



Tasks in time series analysis

- Obtaining data:
  - **Selection** of attributes: Choose relevant indicators
  - Data **collection**
    - Discrete data: Grouping /averaging in time window
    - Continuous data: *Importance of sampling frequency*

- Preprocessing data
  - **Clean data** : Missing data, outliers
  - **Normalization of data** $\frac{X_t - \mu}{\sigma}$; $\frac{X_t - median}{iq}$; $\frac{2X_t - (X_{max} + X_{min})}{X_{max} - X_{min}}$
  - **Eliminate trends /seasonality**: Handle a-priori info explicit / Stationary data. $X_t - X_{t-1}$; $\frac{X_t - X_{t-1}}{X_{t-1}}$; $\log \frac{X_t}{X_{t-1}}$

Parametric / non-parametric data analysis

- Parametric
  - Formulate (restrictive) hypothesis dependent on a set of parameters
  - Find parameters by data-driven optimization [training set]
    - Sensitivity analysis
    - Uncertainty in estimated parameters
    - Robustness
  - Validation of models [test set]
- Non-Parametric
  - Consider a family of universal approximants
  - Fix architecture / parameters by data-driven optimization [training set]
    - Sensitivity analysis
    - Robustness
    - Uncertainty
    - Intelligibility
  - Validation of models [test set]

Classical models in time-series

- Consider the time series

$$X_0, X_1, X_2, \ldots X_{t-1}, X_t, \ldots, X_T$$

  - The series exhibits randomness.
  - The process is covariance-stationary when:
    - **Mean** is time independent

$$E[X_t] = \mu$$

    - **Autocovariance** is independent of time-translations

$$E[(X_{t+\tau} - \mu)(X_t - \mu)] = \gamma_\tau$$

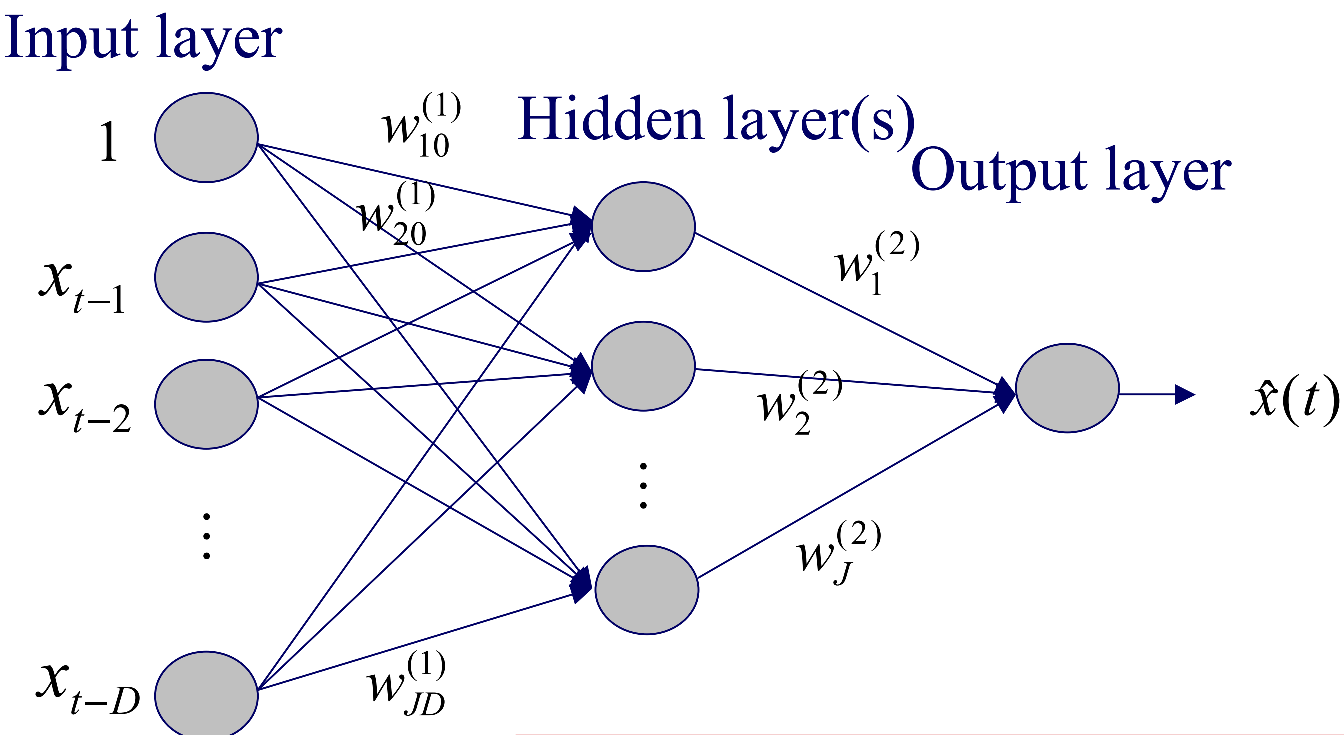
Autorregressive+Moving average models

- Autorregressive model for a time-series

$$X_t = f(\mathbf{X}_t^{[p]}, \mathbf{u}_t^{[q]}; \boldsymbol{\theta}) + u_t$$

  - Vectors of delayed values: $\left[\mathbf{X}_t^{[m]}\right]^+ = [X_{t-1}\ X_{t-2} \cdots X_{t-m}]$
  $\left[\mathbf{u}_t^{[m]}\right]^+ = [u_{t-1}\ u_{t-2} \cdots u_{t-m}]$
  - The systematic term $\hat{\mathbf{X}}_t = f(\mathbf{X}_t^{[p]}, \mathbf{u}_t^{[q]}; \boldsymbol{\theta})$ reflects **trends**.
  - The innovations u_t are **uncorrelated noise**.
  - **Maximization** of the likelihood function yields estimates of the **model parameters.**

Autoregressive (feedforward) MLP



$$\hat{x}_t = \sum_{j=1}^{J} w_j^{(2)} \left(f\left(\sum_{d=1}^{D} w_{jd}^{(1)} x_{t-d} + w_{j0}^{(1)}; \theta_j \right) - c_j \right);$$

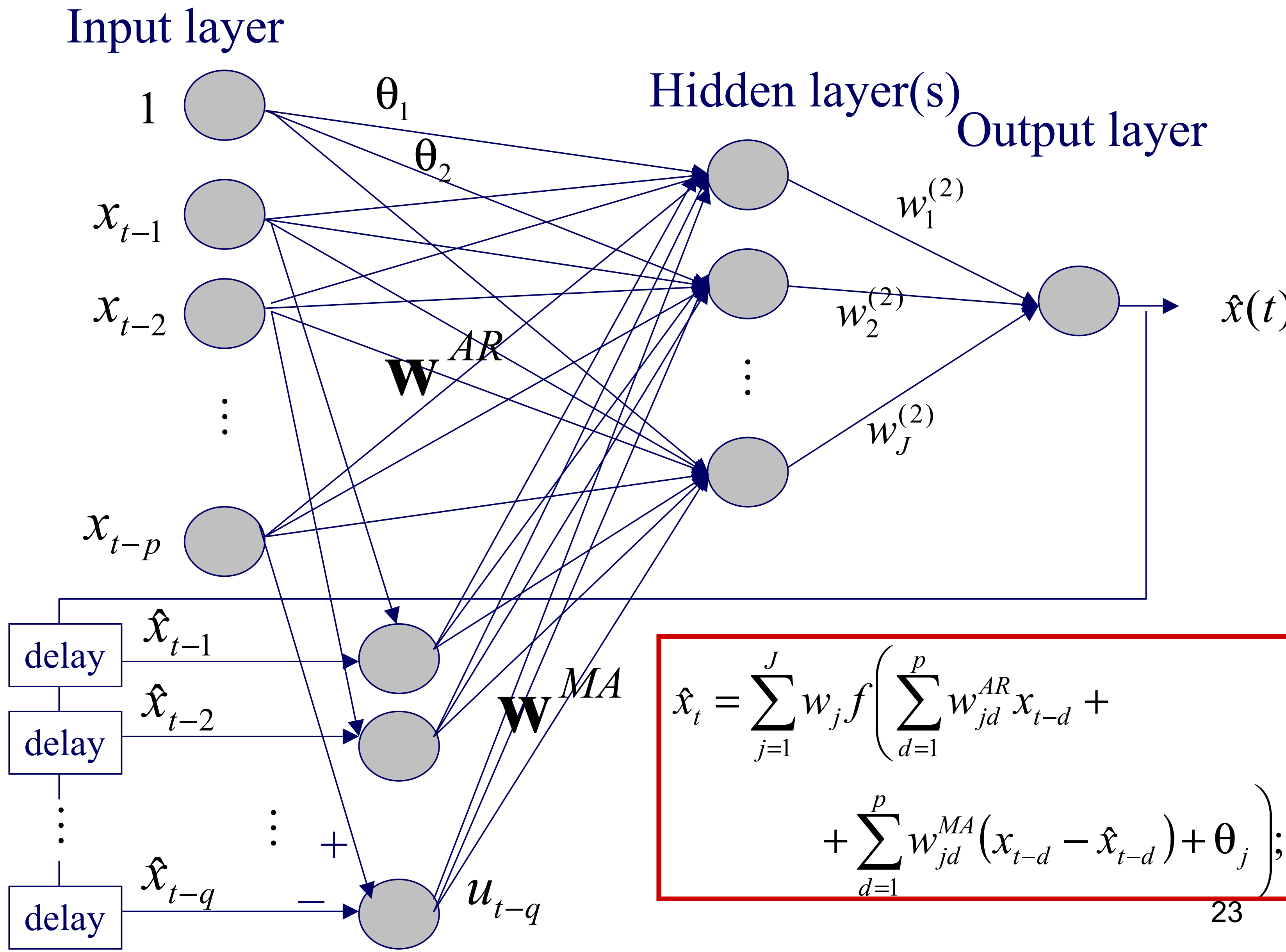
Sigmoidal (logistic)

$$f(x) = \frac{1}{1 - e^{-x}}$$

Hyperbolic tangent:

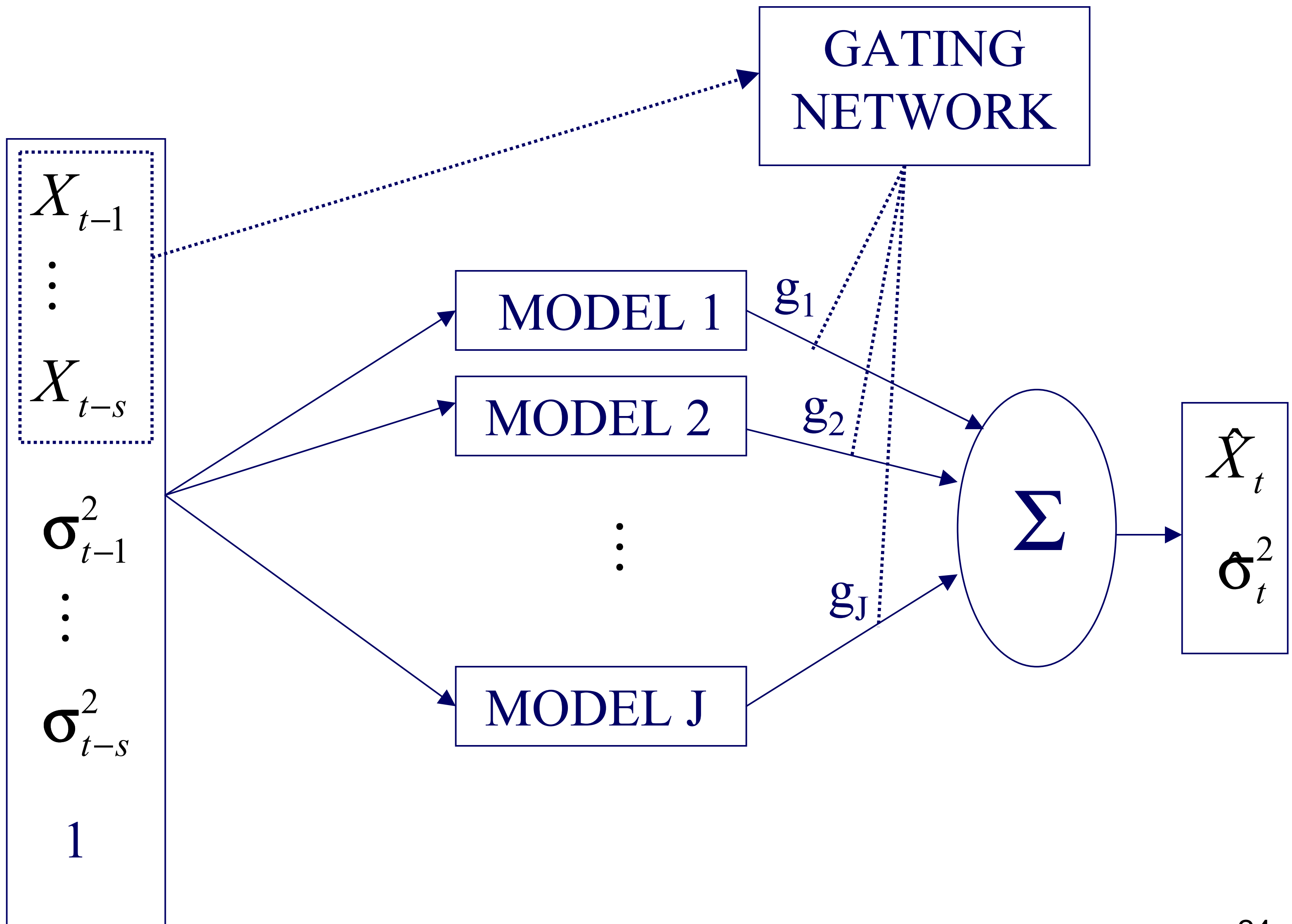
$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

ARMA(p,q) MLP



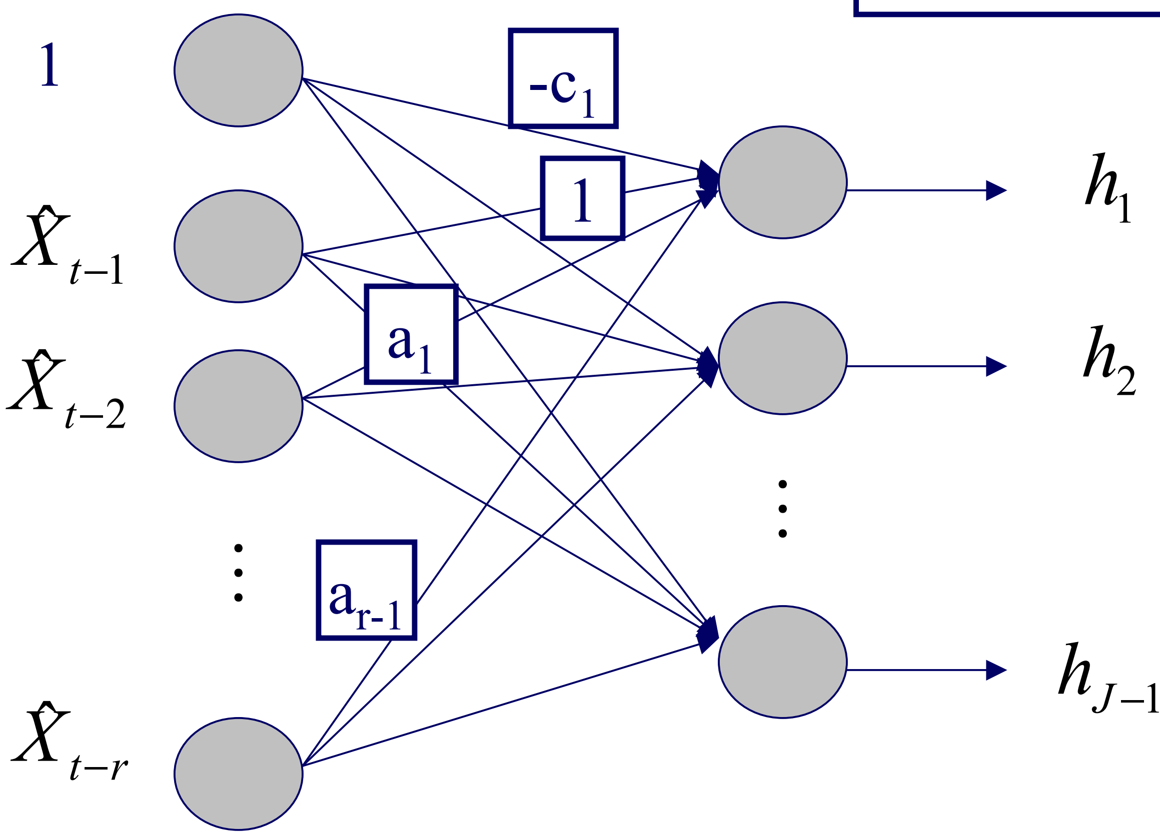
$$\hat{x}_t = \sum_{j=1}^{J} w_j f\left(\sum_{d=1}^{p} w_{jd}^{AR} x_{t-d} + \right.$$
$$\left. + \sum_{d=1}^{p} w_{jd}^{MA} \left(x_{t-d} - \hat{x}_{t-d} \right) + \theta_j \right);$$

Mixture model



Gating Network

$$h_i = \exp\left[b_i \left(\hat{X}_{t-1} + \sum_{k=1}^{r-1} a_i \hat{X}_{t-k-1} - c_i \right) \right]$$

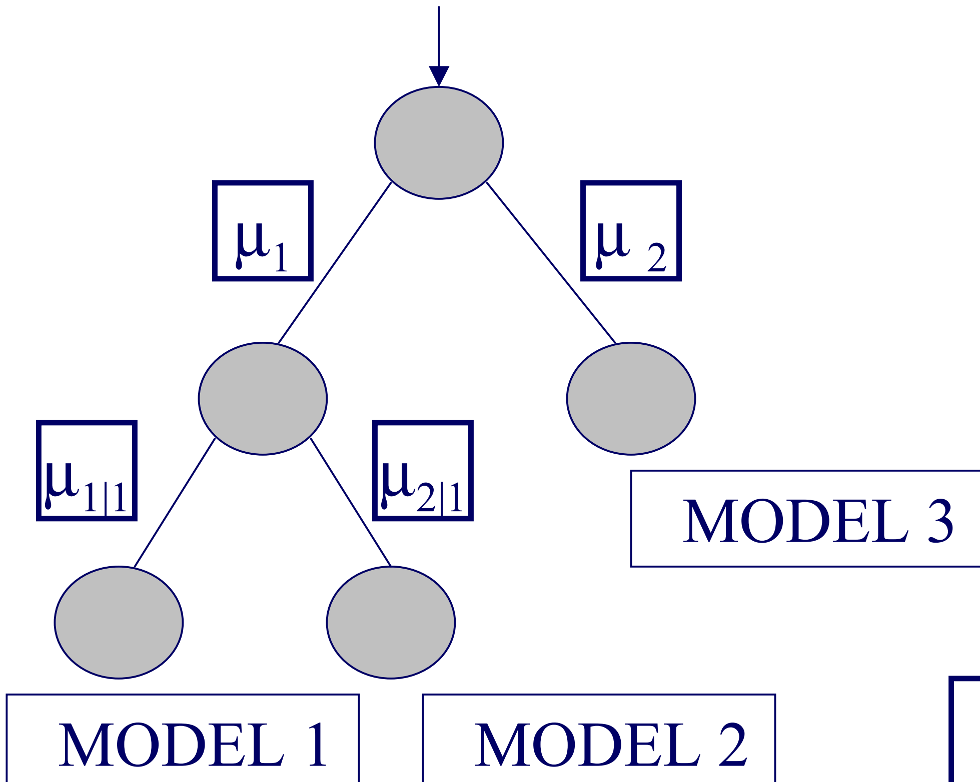


Probabilities

$$g_i = \frac{h_i}{1 + \sum_{j=1}^{J-1} h_j}; \quad i = 1,2,\ldots(J-1) \quad g_J = 1 - \sum_{j=1}^{J-1} g_j$$

Hierarchical mixtures

Input = Vector of Delayed values



$$\mu_1 = \frac{\exp\left[b_1 \left(X_{t-1} + \sum_{k=1}^{r-1} a_{1k} X_{t-k-1} - c_1 \right) \right]}{1 + \exp\left[b_1 \left(X_{t-1} + \sum_{k=1}^{r-1} a_{1k} X_{t-k-1} - c_1 \right) \right]}$$

$$\mu_2 = 1 - \mu_1$$

$$\mu_{1|1} = \frac{\exp\left[b_2 \left(X_{t-1} + \sum_{k=1}^{r-1} a_{2k} X_{t-k-1} - c_2 \right) \right]}{1 + \exp\left[b_2 \left(X_{t-1} + \sum_{k=1}^{r-1} a_{2k} X_{t-k-1} - c_2 \right) \right]}$$

$$\mu_{2|1} = 1 - \mu_{1|1}$$

Model 1 $\mu_{11} = \mu_{1|1}\, \mu_1$;

Model 2 $\mu_{12} = \mu_{2|1}\, \mu_1$;

Model 3 μ_2

Mixture of Gaussians for t-independent pdf

- **Empirical sample** $X_1, X_2, \ldots X_N$
- **Model pdf**

$$P(x) = \sum_{k=1}^{K} p_k \, \mathrm{N}(x; \mu_k, \sigma_k)$$

- **Two steps:**
  - Toss a K-sided loaded dice to choose component.
  - Extract value from the selected model.
- **Advantages:**
  - Close to the normal world.
  - Accounts for leptokurtosis of empirical unconditional distributions in finance.

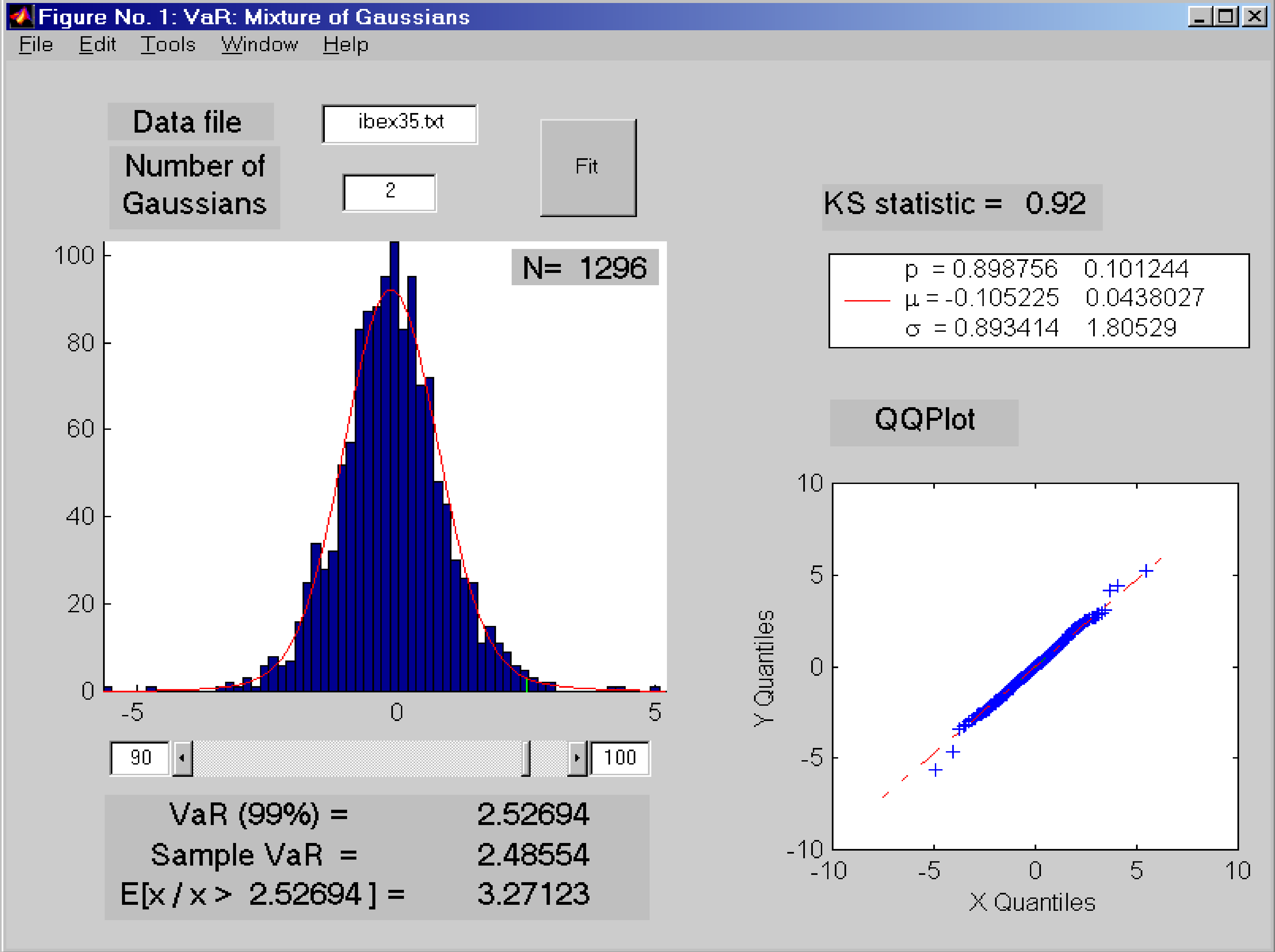


Figure No. 1: VaR: Mixture of Gaussians
File Edit Tools Window Help
Data file
ibex35.txt
Number of Gaussians
2
Fit
KS statistic = 0.92
p = 0.898756 0.101244
μ = -0.105225 0.0438027
σ = 0.893414 1.80529
N= 1296
100
80
60
40
20
0
-5
0
5
90
100
QQPlot
10
5
0
-5
-10
-10
-5
0
5
10
Y Quantiles
X Quantiles
VaR (99%) = 2.52694
Sample VaR = 2.48554
E[x / x > 2.52694] = 3.27123

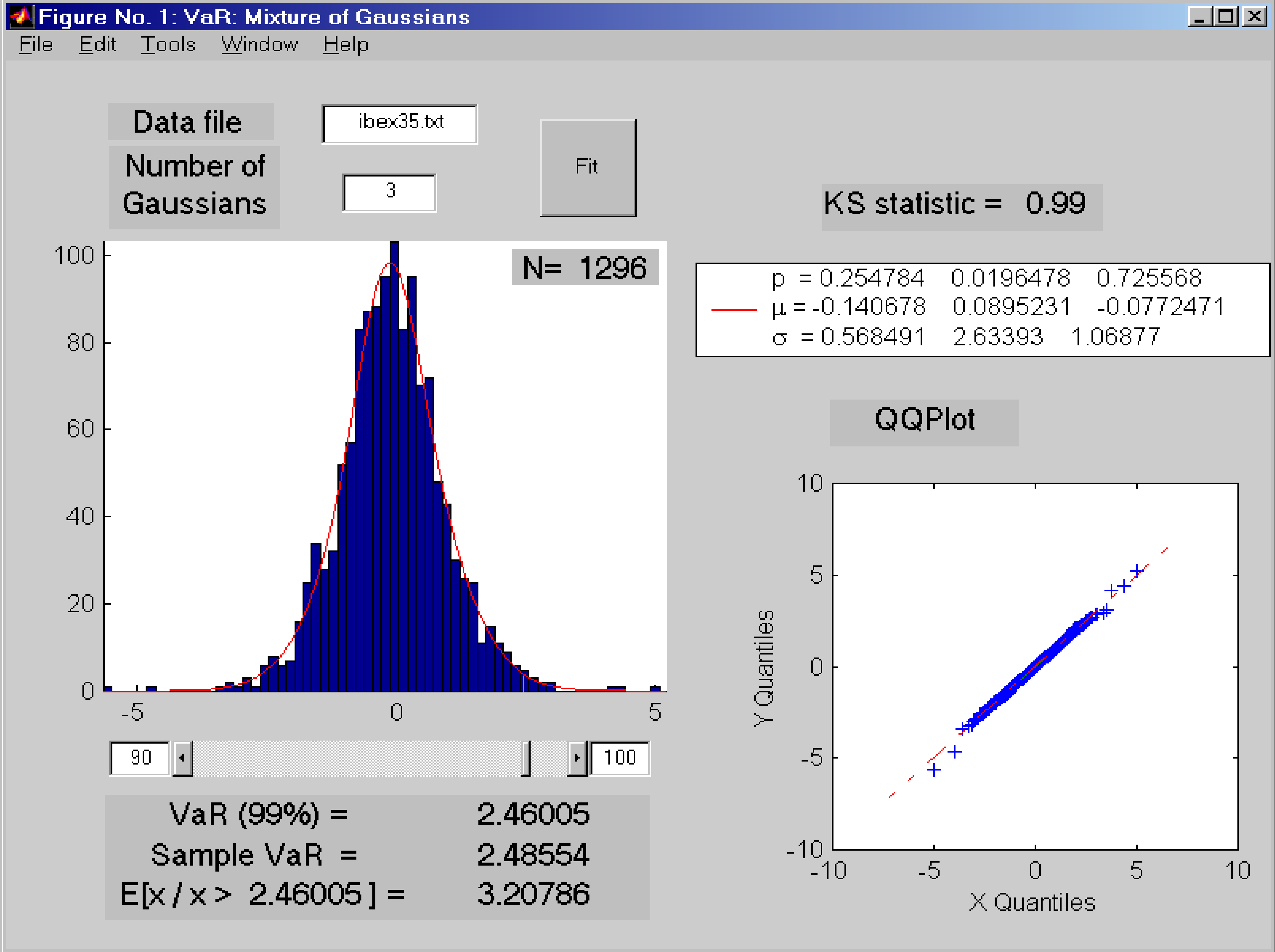


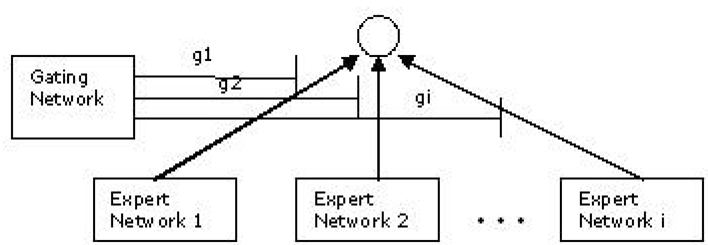
Figure No. 1: VaR: Mixture of Gaussians
File Edit Tools Window Help
Data file
ibex35.txt
Number of Gaussians
3
Fit
KS statistic = 0.99
N= 1296
p = 0.254784 0.0196478 0.725568
μ = -0.140678 0.0895231 -0.0772471
σ = 0.568491 2.63393 1.06877
QQPlot
Y Quantiles
X Quantiles
90
100
VaR (99%) = 2.46005
Sample VaR = 2.48554
E[x / x > 2.46005] = 3.20786

Mixture of Gaussians

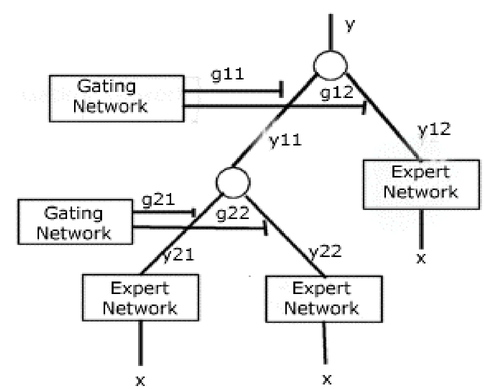
- **Intuition:**
  - Implicitly market forecasts are made in terms of scenarios.
  - Each of these scenarios is characterized by an expected return and a volatility.
  - Markets assign a different probability to each scenario.
- **Dynamical picture?**
  - Direct time aggregation of the process yields a normal model (by Central Limit Theorem).
  - It is possible to construct a discontinuous jump process maintaining the mixture form. Not realistic.

Mixture of AR processes

- **Mixtures of Gaussians + autorregressive dynamics**
  - **In**: Vector of delays (Used in gating network + AR models)
  - **Out**: Next value in time series



No hierarchy

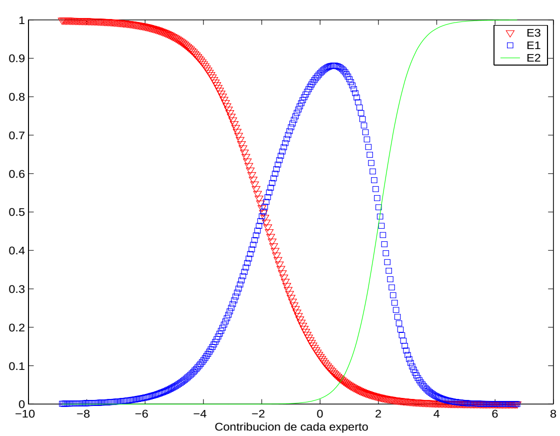


Tree hierarchy

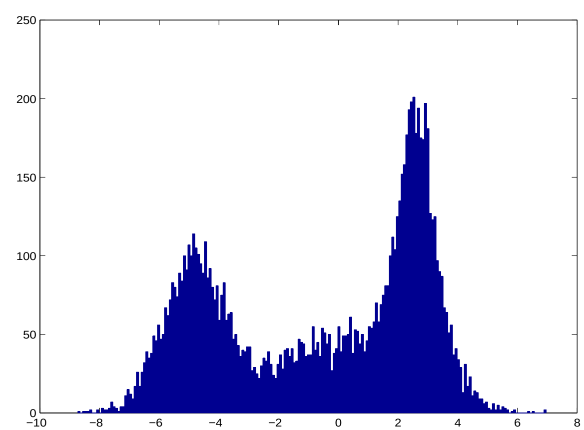
Synthetic data: Example 1

Time series generated by a hierarchical mixture of 3 AR(1) experts

- Expert contributions



- Histogram (unconditional pdf)

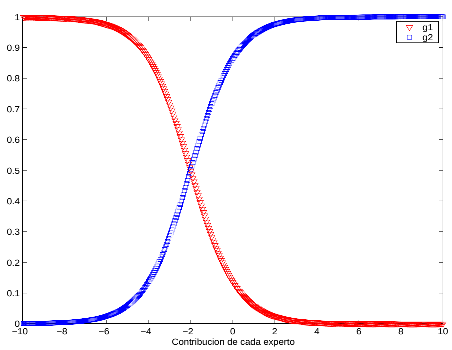


Model 1 fit

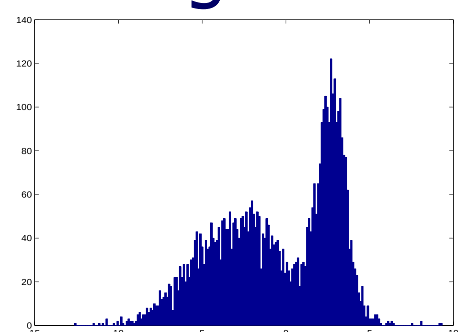
LL Train	LL Test	K-S Test	ECM Test
-17967	-18009	0	0.4645

- Fitting to a mixture of 2 AR(1) experts *(wrong type of model!)*

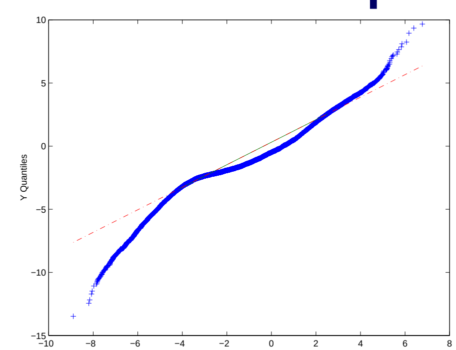
Contributions



Histogram



Percentile plot

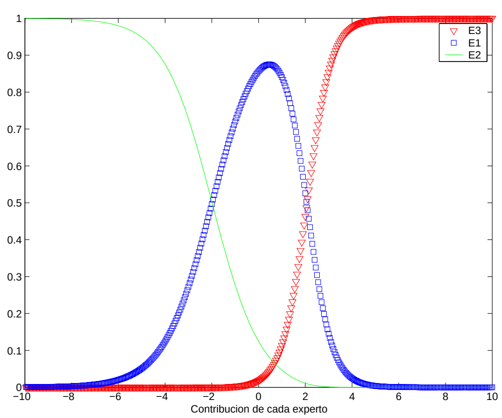


Model 2 fit

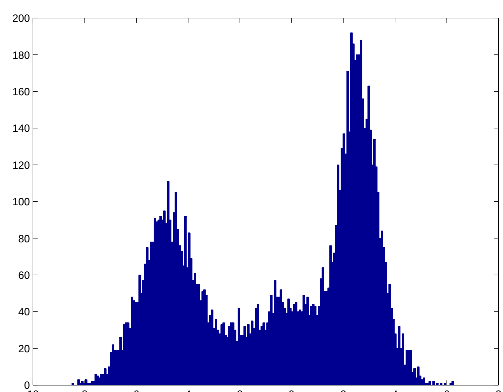
LL Train	LL Test	K-S Test	ECM Test
-16675	-16755	0.9666	0.3164

- Fitting to a mixture of 3 AR(1) experts *(learnable model)*

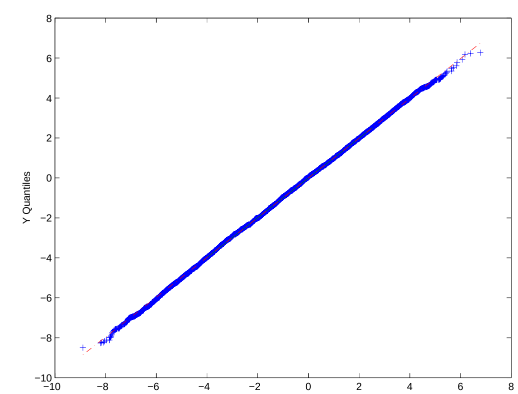
Contributions



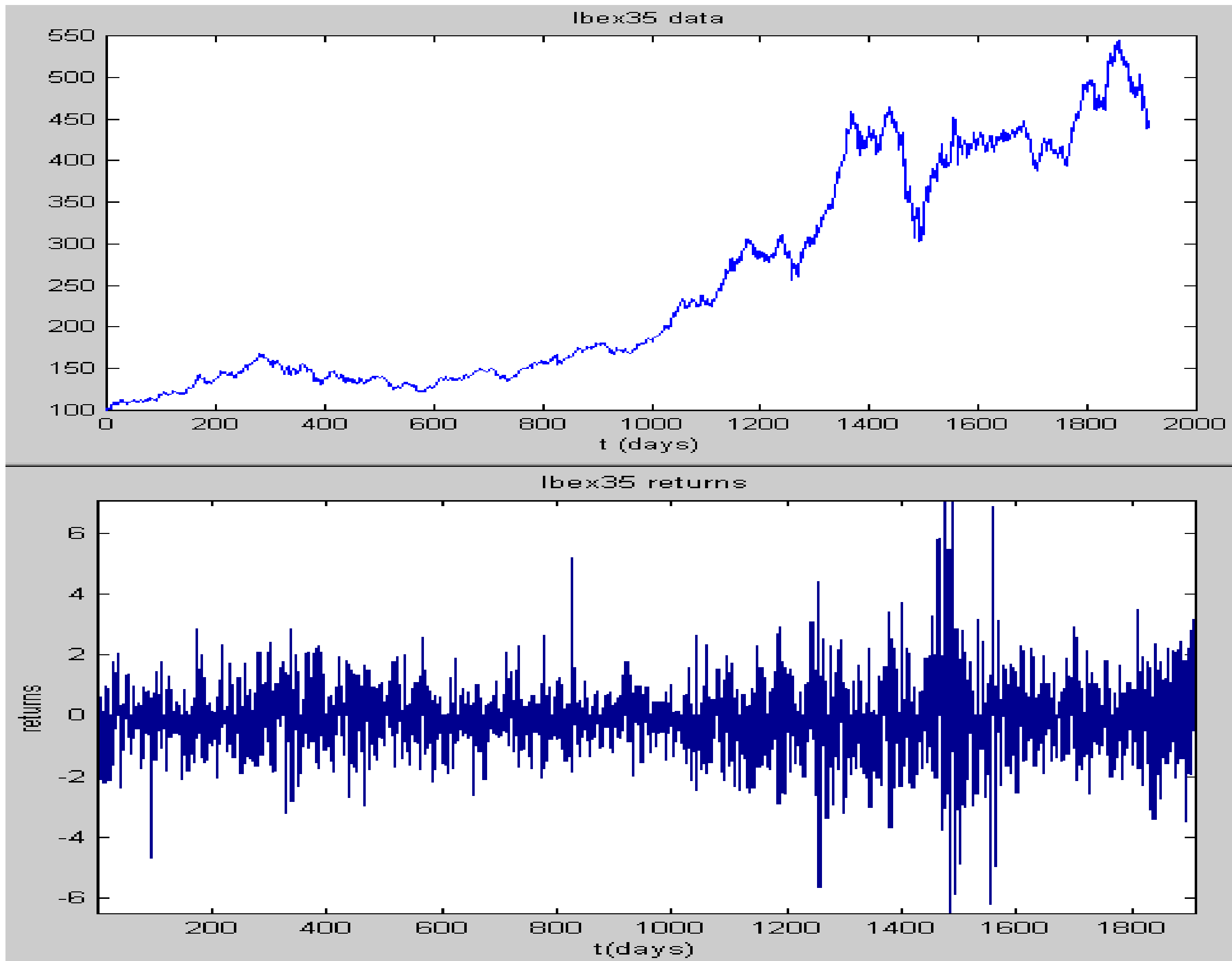
Histogram



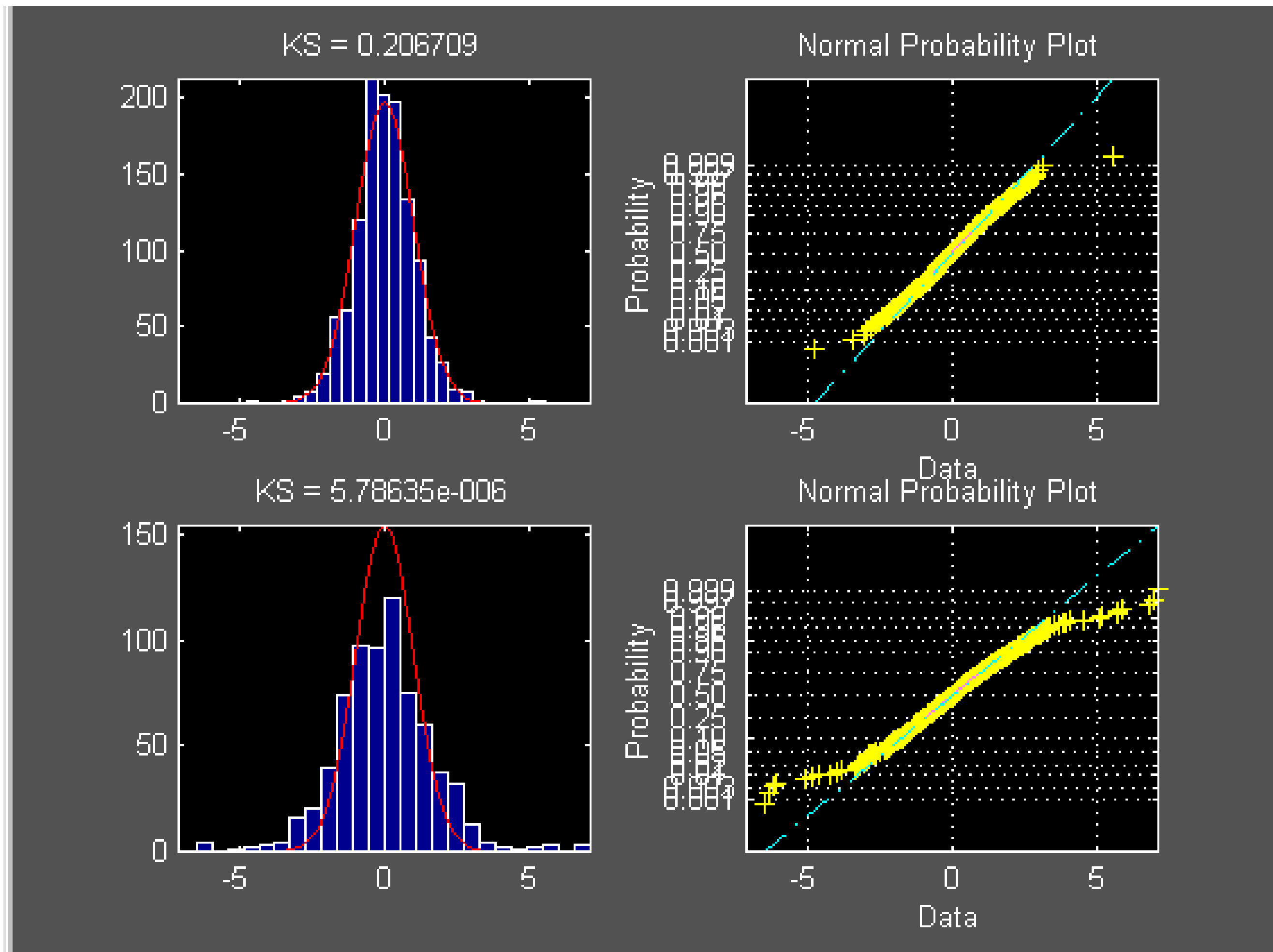
Percentile plot



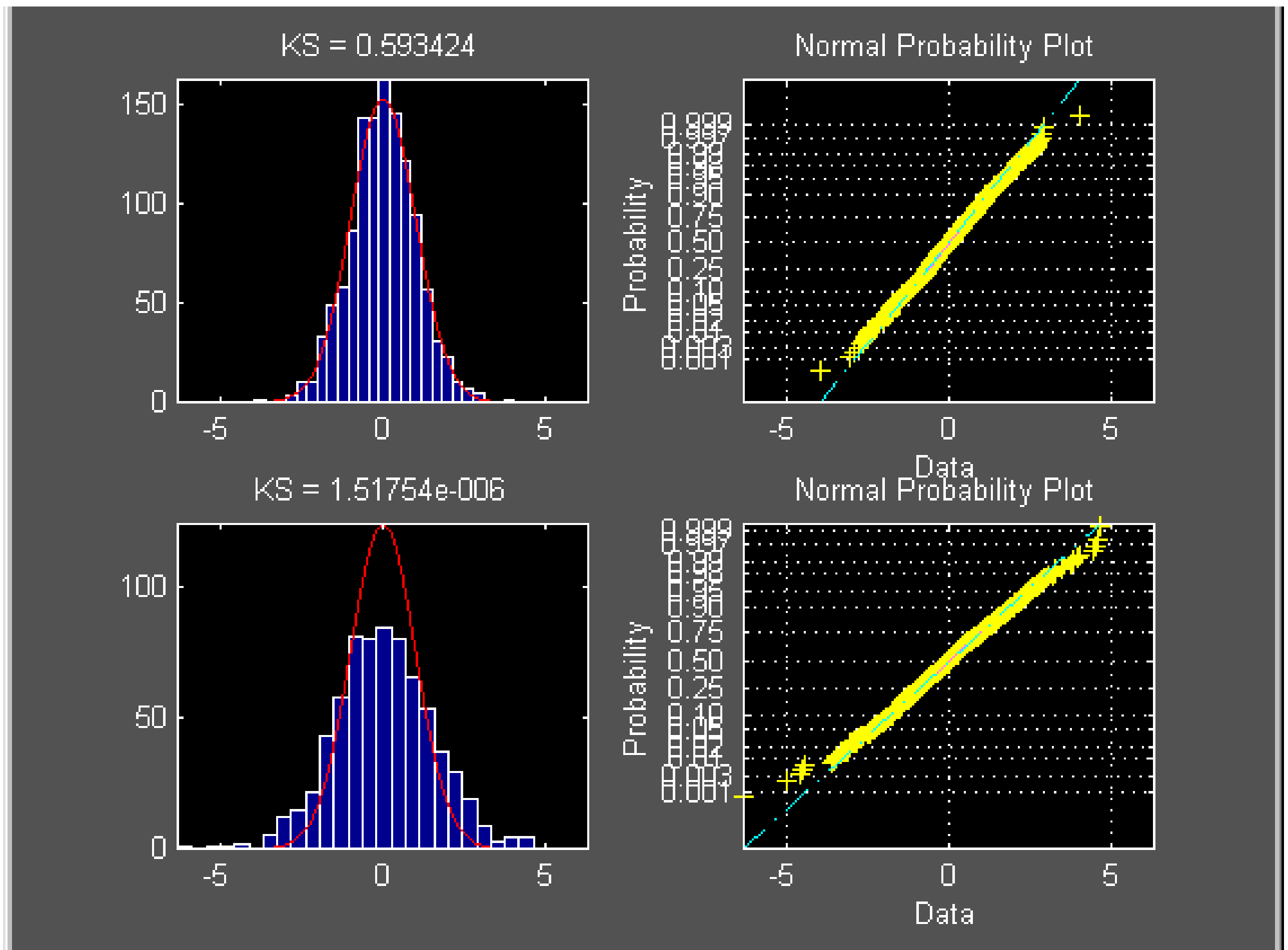
AR(1) fit for Ibex35 (1200 +712 days)



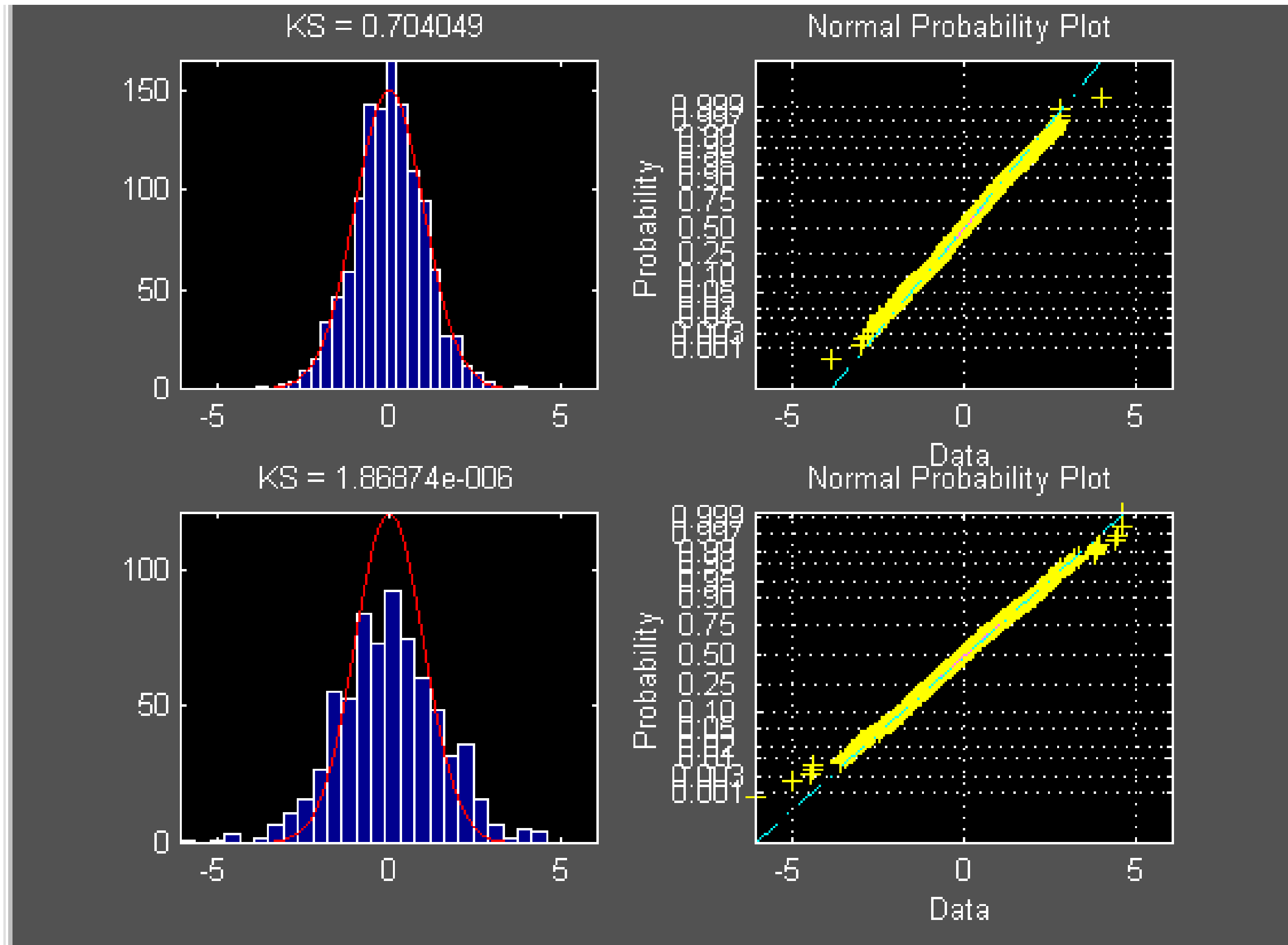
AR(1) fit for Ibex35 (1200 +712 days)



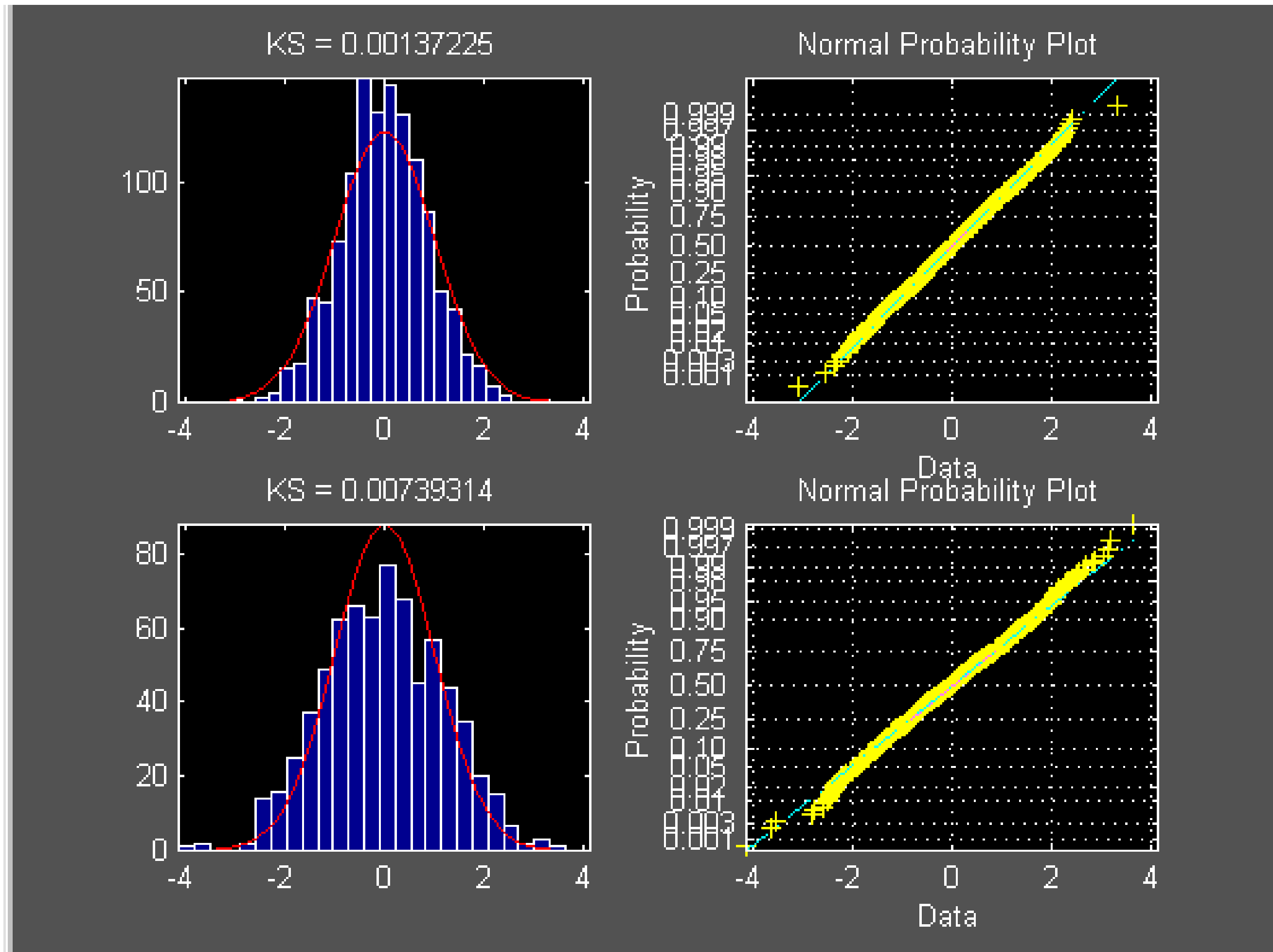
MIX 2 AR(1) fit for Ibex35



MIX 3 AR(1) fit for Ibex35



Hierarchical MIX 3 AR(1) fit for Ibex35



Conclusions and perspectives

- Mixtures of AR(1) models **improve** the results of single AR(1) models in financial returns time series.
- Mixtures of **2 / 3 experts** seem to be **sufficient** to model leptokurtosis and dynamics.
- The introduction of **hierarchy** in the structure of the mixture may significantly improve statistical description of financial time series data.
- To do:
  - Heteroskedasticity
  - Calibration of models to market

Mixture of ARCH processes

- MixARCH $X_t = \phi_{[i]}^{+} \cdot \mathbf{X}_t^{[m]} + u_{[i]}(t),$

  $$\textit{with probability} \;\; g_{[i]}(\mathbf{X}_t^{[r]}, \theta_{[i]})$$

  - The model for the residuals is

    $$u_{[i]}(t) = \sigma_{[i]}(t) Z_t$$

    $$\sigma_{[i]}^2(t) = \kappa_i + \alpha_i^{+} \cdot [u^2]_{[i]}^{[q]}(t)$$

  - The quantities Z_t are assumed to be *N(0,1)*

Mixture of GARCH processes

- MixGARCH $\hat{X}_t = \phi^{+}_{[i]} \cdot \hat{\mathbf{X}}^{[m]}_t + u_{[i]}(t),$

  $$\textit{with probability} \;\; g_{[i]}(\mathbf{X}^{[r]}_t, \theta_{[i]})$$

  - The model for the residuals is

  $$u_{[i]}(t) = \sigma_{[i]}(t) Z_t$$

  $$\sigma^2_{[i]}(t) = \kappa_i + \alpha^{+}_i \cdot [u^2]^{[q]}_{[i]}(t) + \beta^{+}_i \cdot [\sigma^2]^{[p]}_{[i]}(t)$$

  - The quantities Z_t are assumed to be *N(0,1)*

AR(1) / ARCH(1) for IBEX35

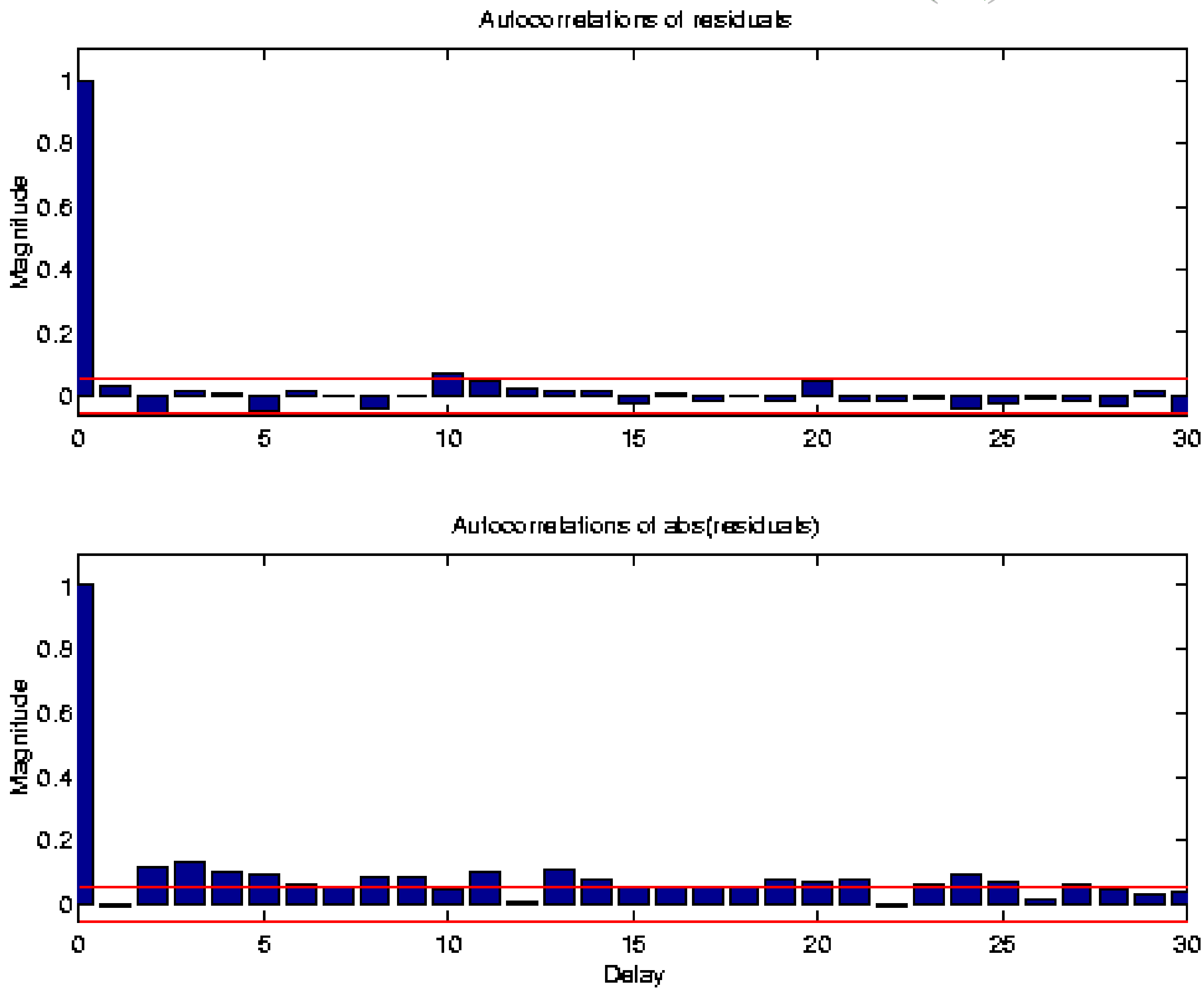
- The maximum-likelihood fit of the time-series IBEX35 yields the model

$$\hat{X}_t = 0.1129\hat{X}_{t-1} + \sigma_t Z_t$$

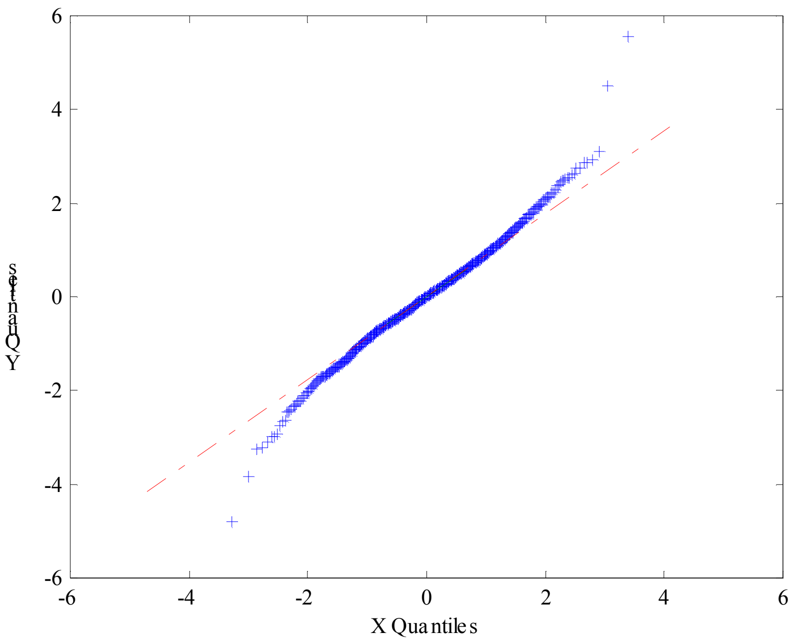
$$\sigma_t^2 = 0.9097 + 0.1118\left(\hat{X}_{t-1} - 0.1129\hat{X}_{t-2}\right)^2$$

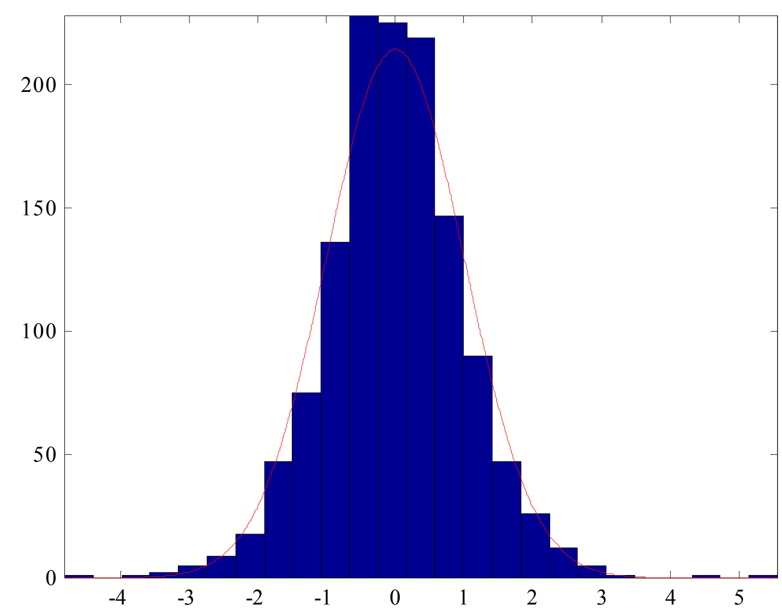
  - The quantities Z_t are assumed to follow a N(0,1) distribution.

Residual correlations: ARCH(1)



Normality hypothesis: ARCH(1)





KS Test = 0.12

MIXARCH for IBEX35

- The mixture model is

$$\text{Model 1} \quad \hat{X}_t = 0.0559\hat{X}_{t-1} + \sigma_t Z_t$$

$$\sigma_t^2 = 2.2194 + 0.1976\left(\hat{X}_{t-1} - 0.0559\hat{X}_{t-2}\right)^2$$

$$\text{Model 2} \quad \hat{X}_t = 0.1380\hat{X}_{t-1} + \sigma_t Z_t$$

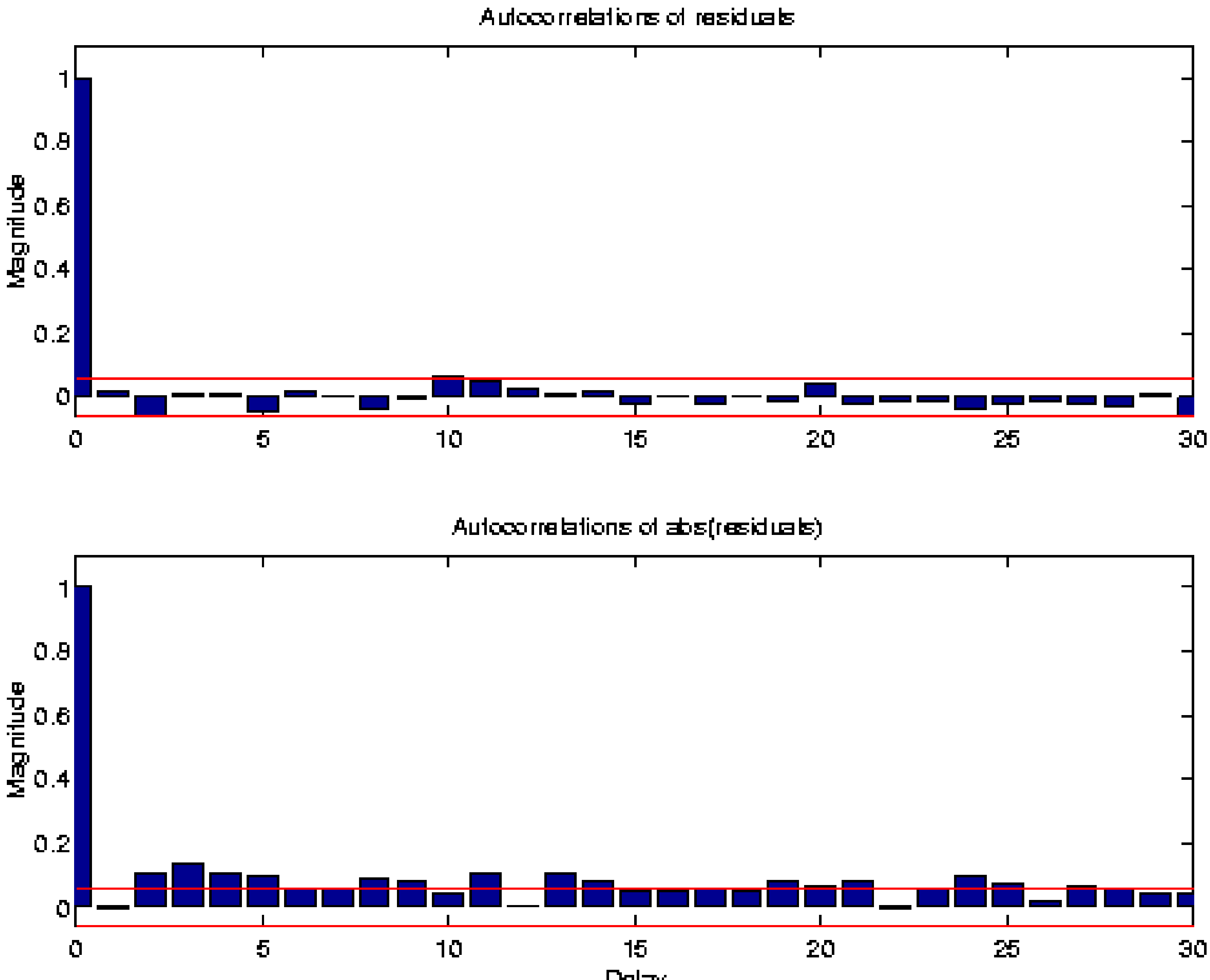
$$\sigma_t^2 = 0.6820 + 0.03821\left(\hat{X}_{t-1} - 0.1380\hat{X}_{t-2}\right)^2$$

  - The probabilities for the mixture are

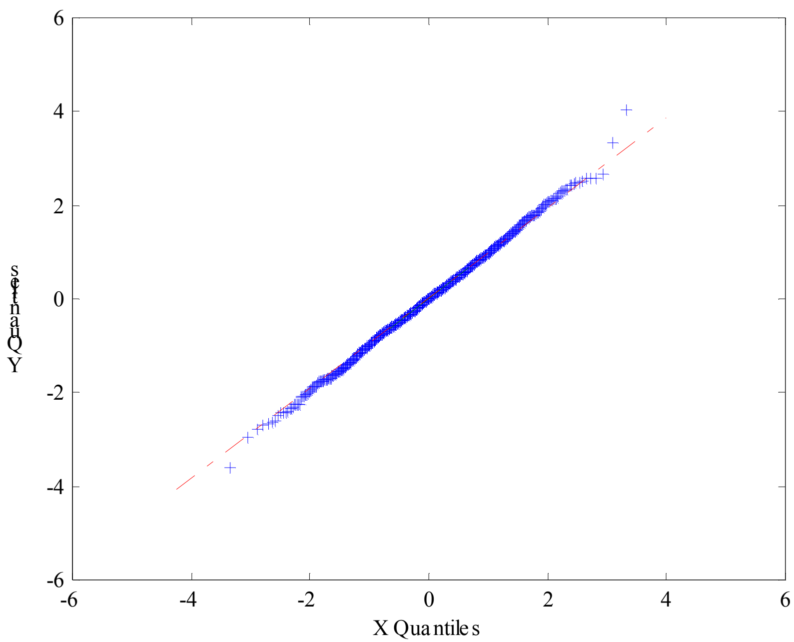
$$g_{[1]}(X_{t-1}) = \frac{1}{1 + \exp\{-0.6839(X_{t-1} - 2.5155)\}};$$

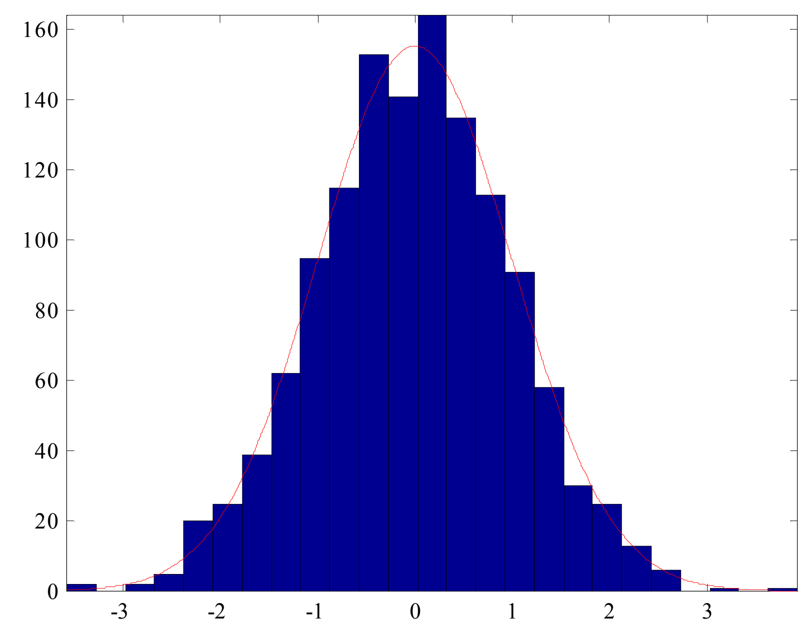
$$g_{[2]}(X_{t-1}) = 1 - g_{[1]}(X_{t-1})$$

Residual correlations: MIXARCH



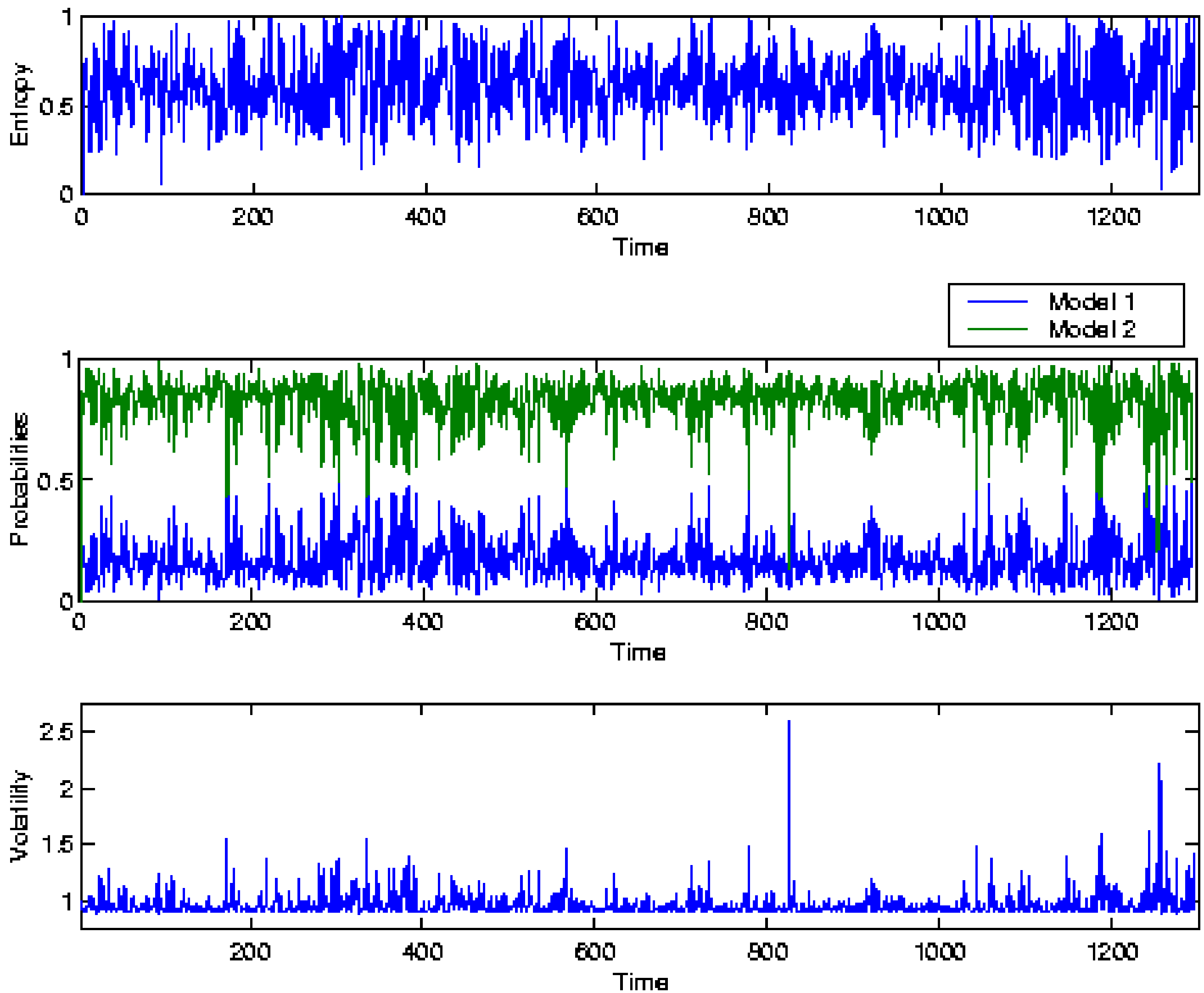
Normality hypothesis: MixARCH(1)





KS Test = 0.83

MIXARCH Model fit



AR(1) / GARCH(1,1) for IBEX35

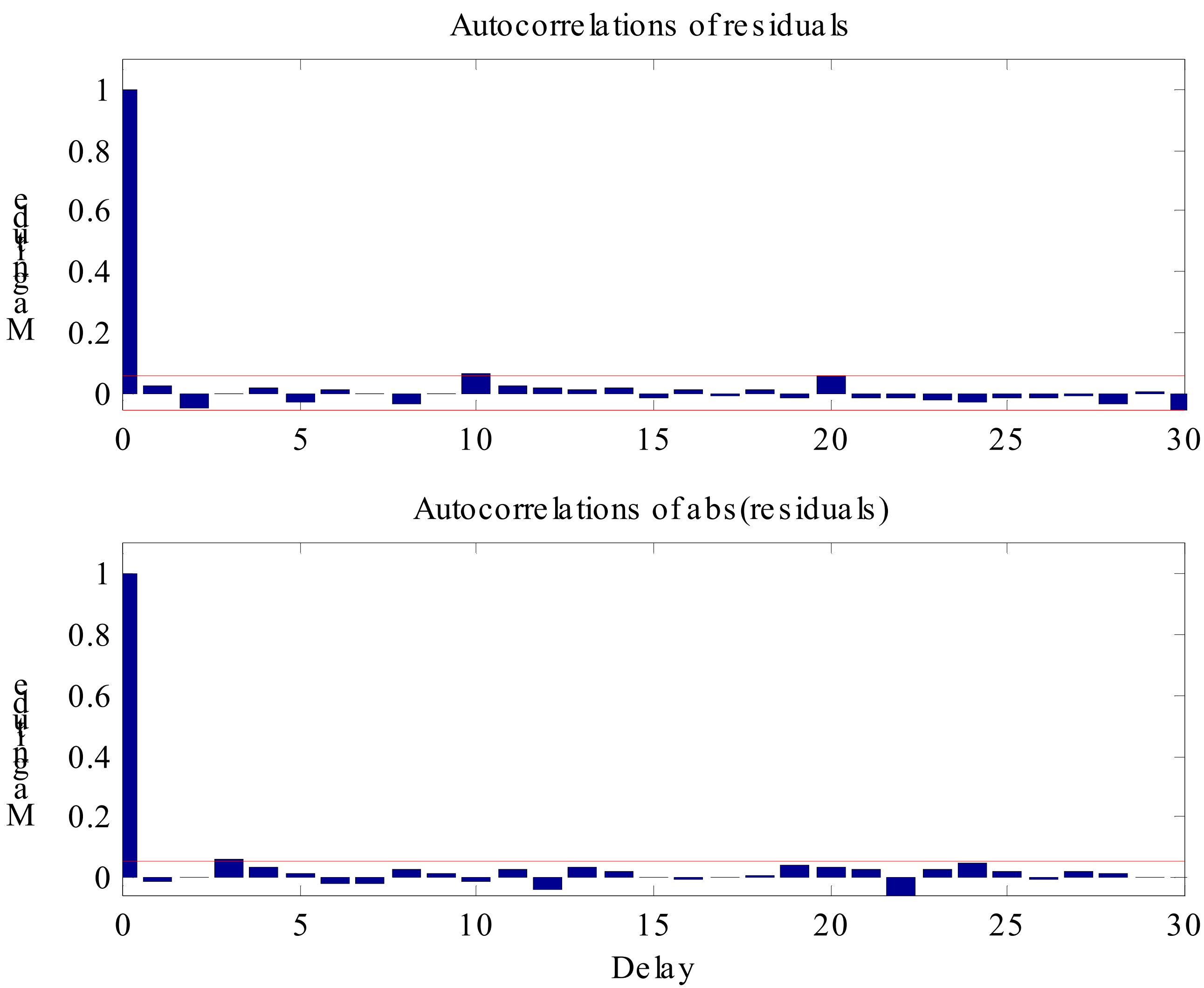
- The maximum-likelihood fit of the time-series IBEX35 yields the model

$$\hat{X}_t = 0.1358\hat{X}_{t-1} + \sigma_t Z_t$$

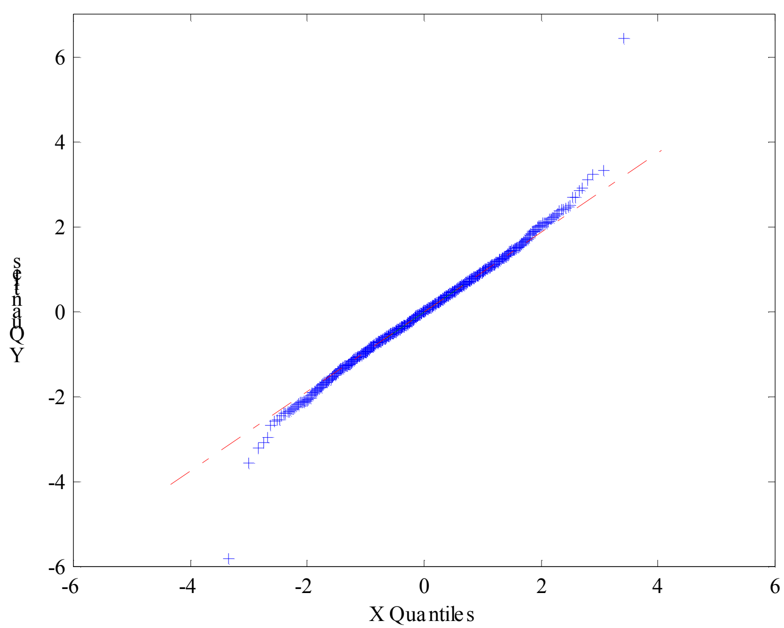
$$\sigma_t^2 = 0.0527 + 0.0755\left(\hat{X}_{t-1} - 0.1358\hat{X}_{t-2}\right)^2 + 0.8733\sigma_{t-1}^2$$

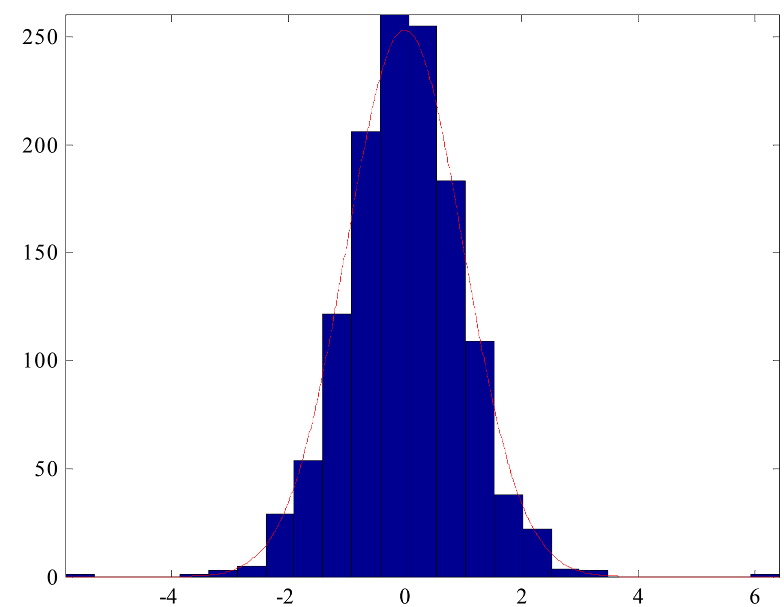
  - The quantities Z_t are assumed to follow a N(0,1) distribution.

Residual correlations: GARCH



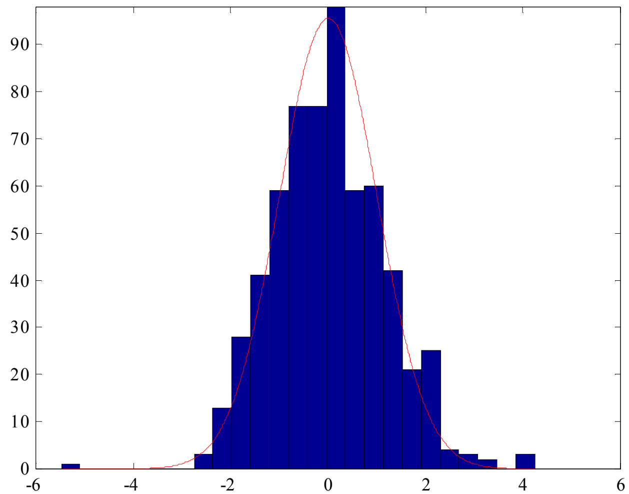
Normality hypothesis: GARCH(1,1)



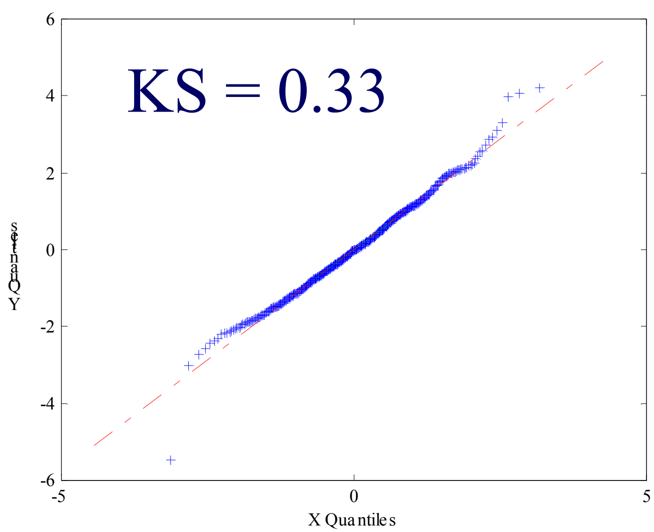


KS Test = 0.56

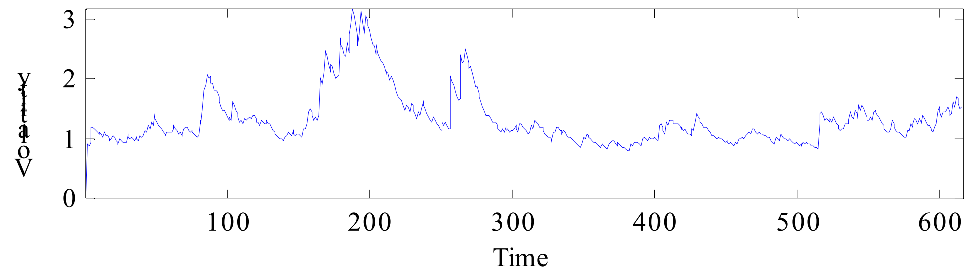
Test Data



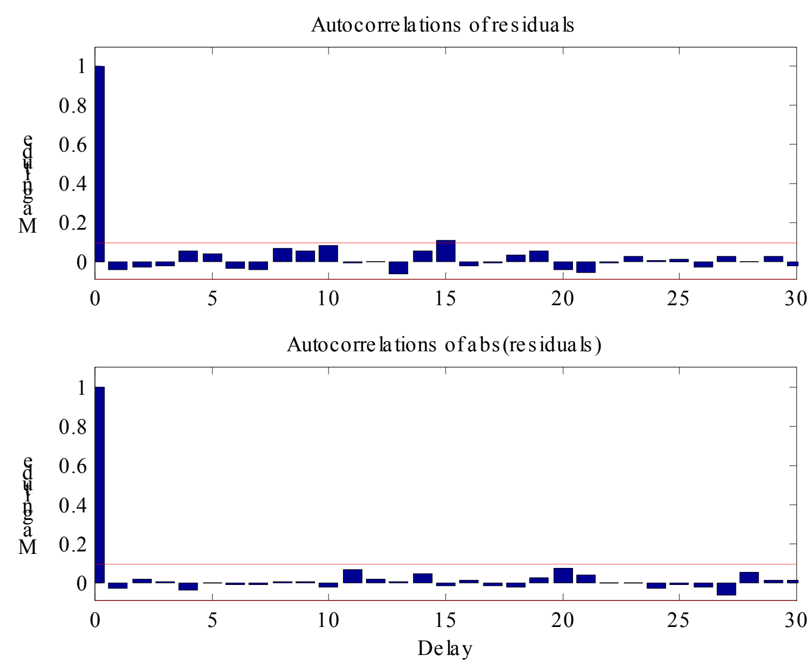
90
80
70
60
50
40
30
20
10
0
-6
-4
-2
0
2
4
6



KS = 0.33
X Quantiles
6
4
2
0
-2
-4
-6
-5
0
5



Time
3
2
1
0
100
200
300
400
500
600



Autocorrelations of residuals
Autocorrelations of abs(residuals)
Delay
1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30

MIXGARCH for IBEX35

- The mixture model is

Model 1 $\hat{X}_t = 0.1255\hat{X}_{t-1} + \sigma_t Z_t$

$$\sigma_t^2 = 0.0156 + 0.0778\left(\hat{X}_{t-1} - 0.1255X_{t-2}\right)^2 + 0.8937\,\sigma_{t-1}^2$$

Model 2 $\hat{X}_t = 0.3314\hat{X}_{t-1} + \sigma_t Z_t$

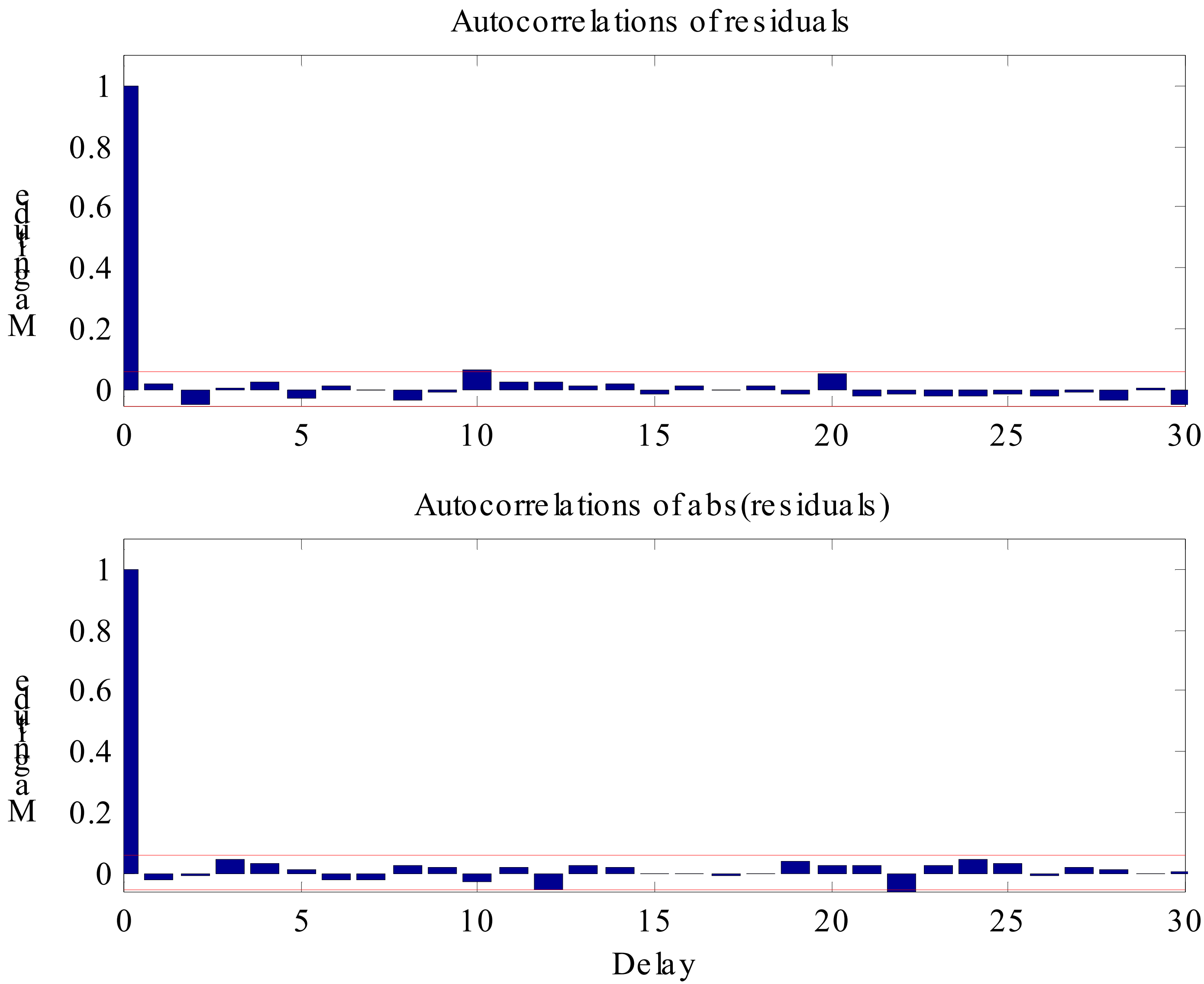
$$\sigma_t^2 = 2.6230 + 0.0000\left(\hat{X}_{t-1} - 0.3314\hat{X}_{t-2}\right)^2 + 0.0285\,\sigma_{t-1}^2$$

- The probabilities for the mixture are

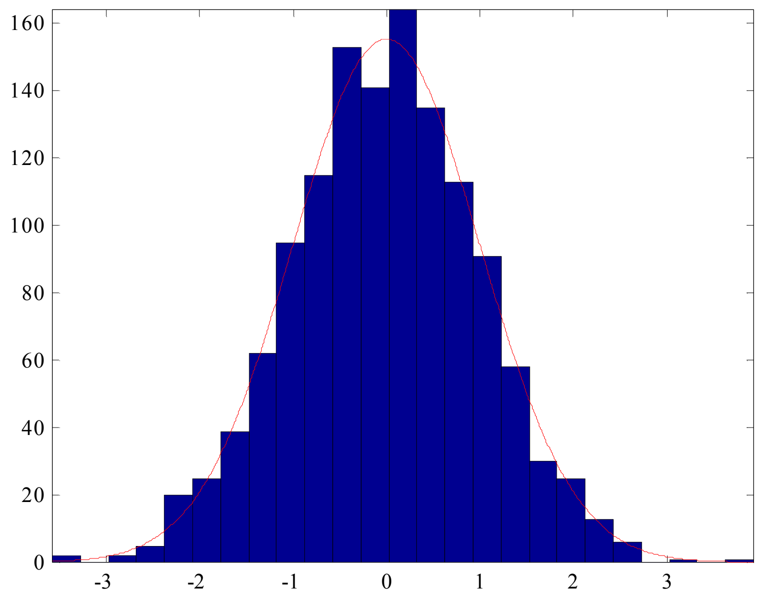
$$g_{[1]}(X_{t-1}) = \frac{1}{1+\exp\left\{0.5418\,(\hat{X}_{t-1} - 4.8710)\right\}};$$

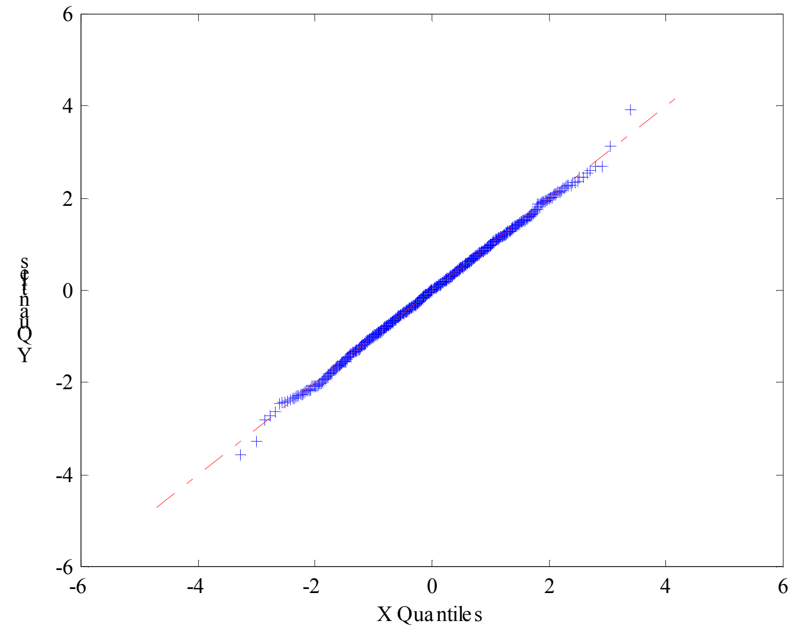
$$g_{[2]}(X_{t-1}) = 1 - g_{[1]}(X_{t-1})$$

Residual correlations: MIXGARCH



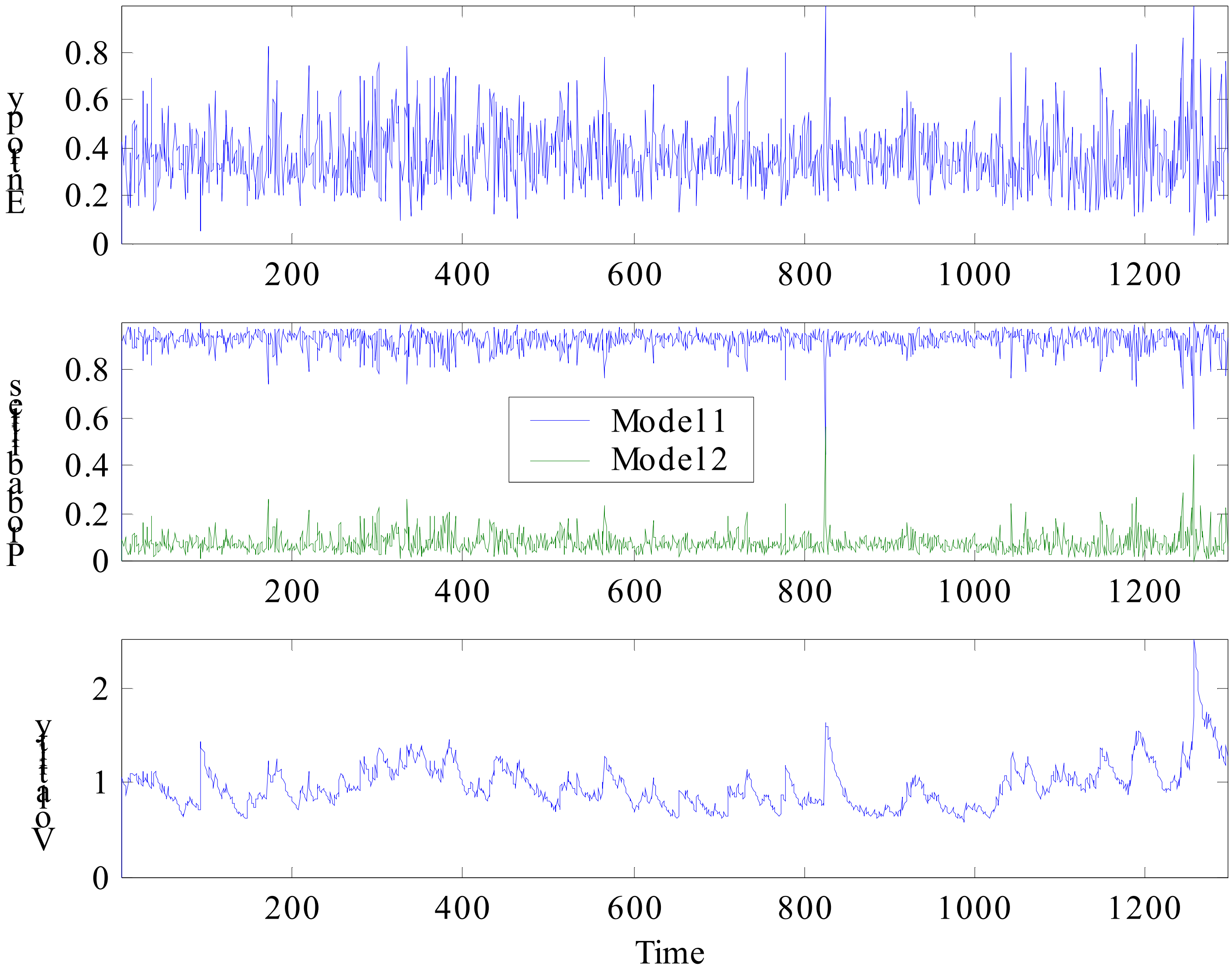
Normality hypothesis: MIXGARCH



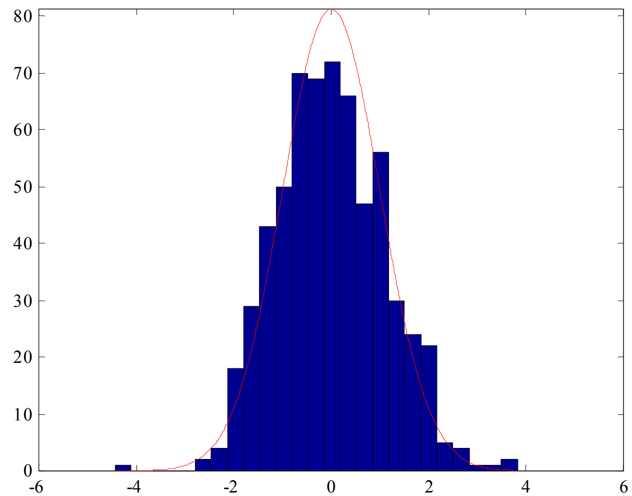


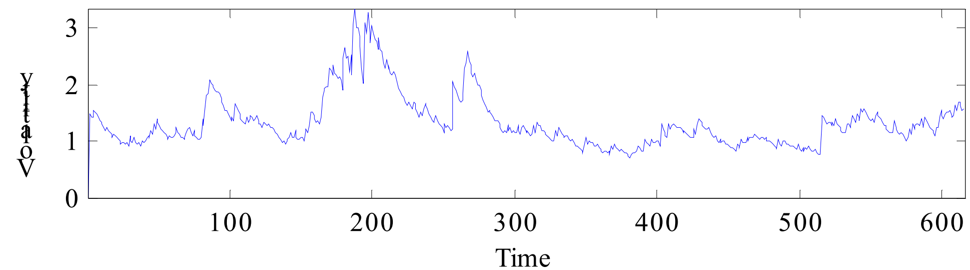
KS test = 0.95

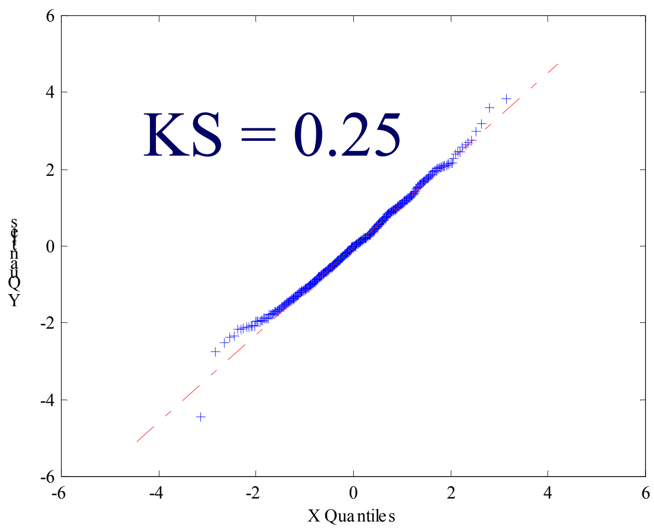
MIXGARCH Model fit

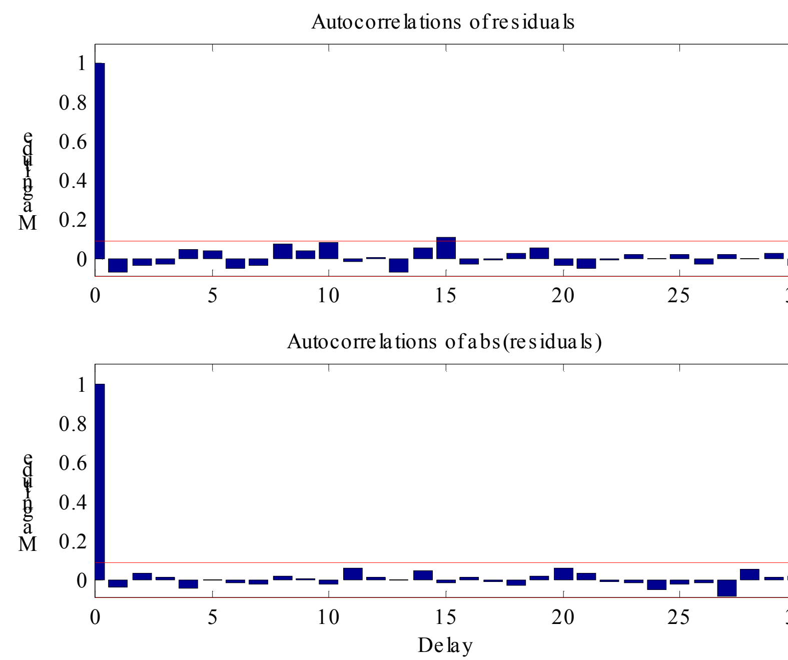


Test Data




Time


KS = 0.25
X Quantiles


Autocorrelations of residuals
Autocorrelations of abs(residuals)
Delay