Preliminary Experiments with On-line Adaptive GARCH Models

Miguel Lázaro-Gredilla,

M. Ortega-Moral, J. Navarro-Abellán, A.R. Figueiras-Vidal

{miguel,ortegam,jnavarro,arfv}@tsc.uc3m.es

Department of Signal Theory and Communications

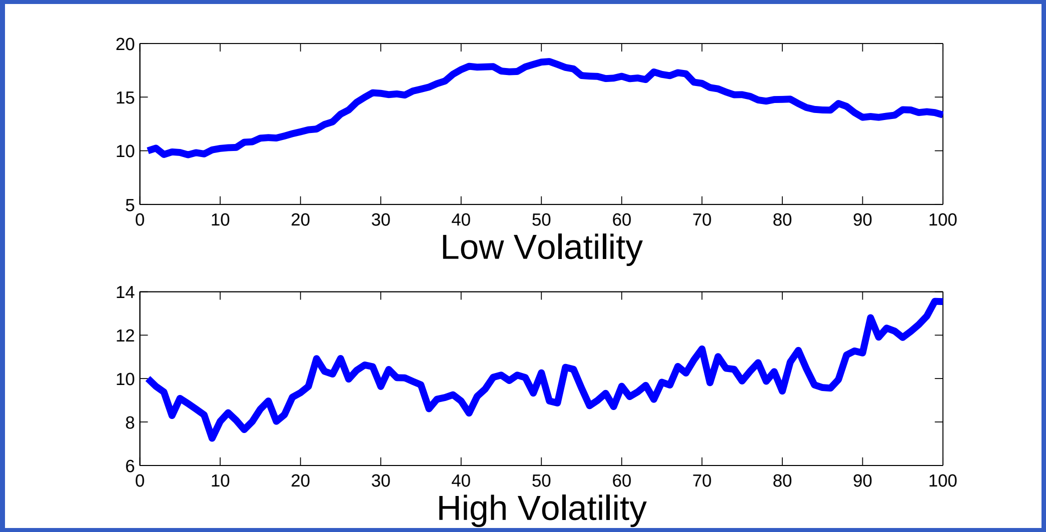
Universidad Carlos III de Madrid

Content

- What is Volatility?
- The Volatility Prediction Problem
- The GARCH Model
- LMS-GARCH(1,1) and RLS-GARCH(1,1) Algorithms
- Experiments
- Conclusions

What is Volatility? (I/II)

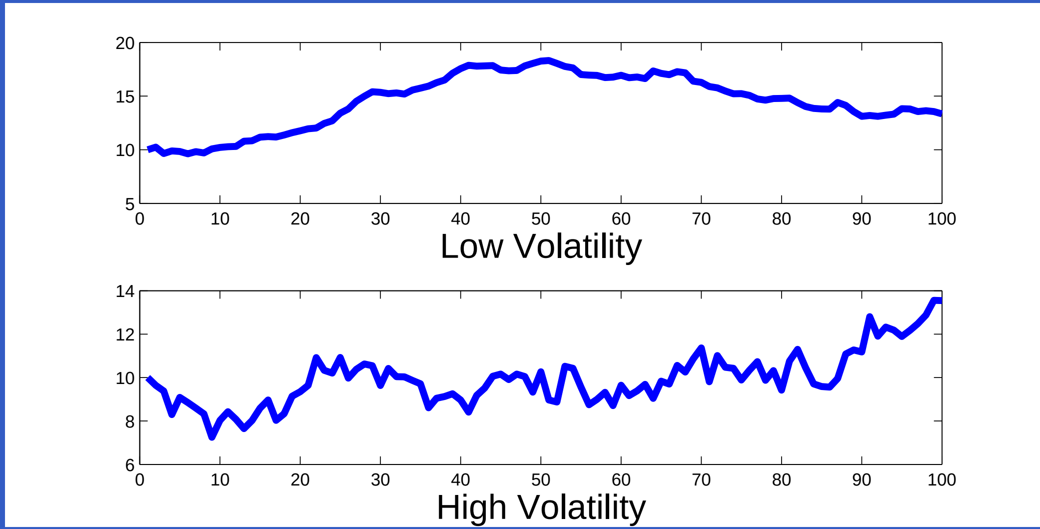
- Intuitive definition: How much can a security price vary?



What is Volatility? (I/II)

Departament of Signal Theory and Communications

- Intuitive definition: How much can a security price vary?



- Formal definition: Standard deviation of the return series, conditional on all the information available on the return process up to the previous time period.

What is Volatility? (II/II)

Departament of Signal Theory and Communications

- Return series:

$$r_t = \log\left(\frac{c_t}{c_{t-1}}\right)$$

What is Volatility? (II/II)

- Return series:

$$r_t = \log\left(\frac{c_t}{c_{t-1}}\right)$$

- Volatility:

$$\sigma_t = \sqrt{Var(r_t|\mathcal{F}_{t-1})}$$

Departament of Signal Theory and Communications

The Volatility Prediction Problem

Departament of Signal Theory
and Communications

- Volatility prediction aims to estimate σ_{t+1} at time t

The Volatility Prediction Problem

Departament of Signal Theory and Communications

- Volatility prediction aims to estimate σ_{t+1} at time t
- However true volatility σ_{t+1} is unobservable, even at time $t+1$!

The Volatility Prediction Problem

Departament of Signal Theory and Communications

- Volatility prediction aims to estimate σ_{t+1} at time t
- However true volatility σ_{t+1} is unobservable, even at time $t+1$!
- This makes difficult even evaluating predictions coming from different models

The Volatility Prediction Problem

Departament of Signal Theory and Communications

- Volatility prediction aims to estimate σ_{t+1} at time t
- However true volatility σ_{t+1} is unobservable, even at time $t+1$!
- This makes difficult even evaluating predictions coming from different models
- Volatility prediction is useful for:
  - Risk management

The Volatility Prediction Problem

Departament of Signal Theory and Communications

- Volatility prediction aims to estimate σ_{t+1} at time t
- However true volatility σ_{t+1} is unobservable, even at time $t+1$!
- This makes difficult even evaluating predictions coming from different models
- Volatility prediction is useful for:
  - Risk management
  - Option pricing

The Volatility Prediction Problem

- Volatility prediction aims to estimate σ_{t+1} at time t
- However true volatility σ_{t+1} is unobservable, even at time $t+1$!
- This makes difficult even evaluating predictions coming from different models
- Volatility prediction is useful for:
  - Risk management
  - Option pricing
  - Trading and hedging strategies

The GARCH Model (I/II)

- GARCH model was proposed by Bollerslev in 1986 as a generalization of Engle's ARCH model (1982). GARCH stands for Generalized AutoRegresive Conditional Heteroskedasticity.

The GARCH Model (I/II)

Departament of Signal Theory and Communications

- GARCH model was proposed by Bollerslev in 1986 as a generalization of Engle's ARCH model (1982). GARCH stands for Generalized AutoRegresive Conditional Heteroskedasticity.
- This model assumes that the return series can be modelled as:

$$r_t = C + \varepsilon_t$$

where $p(\varepsilon_t|\mathcal{F}_{t-1})$ is $N(0, \sigma_t^2)$ and C is the (usually neglected) bias of the distribution.

The GARCH Model (I/II)

Departament of Signal Theory and Communications

- GARCH model was proposed by Bollerslev in 1986 as a generalization of Engle's ARCH model (1982). GARCH stands for Generalized AutoRegresive Conditional Heteroskedasticity.
- This model assumes that the return series can be modelled as:

$$r_t = C + \varepsilon_t$$

where $p(\varepsilon_t|\mathcal{F}_{t-1})$ is $N(0, \sigma_t^2)$ and C is the (usually neglected) bias of the distribution.

- The square of the volatility, σ_t^2, is then modelled as:

$$\hat{\sigma}_t^2 = \alpha_0 + \sum_{i=1}^{Q} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{P} \beta_j \hat{\sigma}_{t-1}^2$$

which is a generic GARCH(P,Q) model. $P = 0$ yields ARCH.

The GARCH Model (II/II)

- GARCH(1,1) models are usually preferred
  - Fewer parameters (Three: α_0, α_1, β_1)
  - Reliable out-of-sample forecasts

The GARCH Model (II/II)

Departament of Signal Theory and Communications

- GARCH(1,1) models are usually preferred
  - Fewer parameters (Three: α_0, α_1, β_1)
  - Reliable out-of-sample forecasts
- Model parameters are selected so that they maximize the LLF

$$\text{LLF} = \log\left(\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} e^{-\frac{r_t^2}{\hat{\sigma}_t^2}}\right) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log\hat{\sigma}_t^2 + \frac{r_t^2}{\hat{\sigma}_t^2}\right)$$

where T is the number of time periods (i. e., days).

The GARCH Model (II/II)

- GARCH(1,1) models are usually preferred
  - Fewer parameters (Three: α_0, α_1, β_1)
  - Reliable out-of-sample forecasts
- Model parameters are selected so that they maximize the LLF

$$\text{LLF} = \log\left(\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} e^{-\frac{r_t^2}{\hat{\sigma}_t^2}}\right) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log \hat{\sigma}_t^2 + \frac{r_t^2}{\hat{\sigma}_t^2}\right)$$

where T is the number of time periods (i. e., days).

- Since $\hat{\sigma}_t^2$ is obtained iteratively, we cannot use gradient search. LLF maximization is carried out through numerical optimization, evaluating the LLF at different points $\{\alpha_i, \beta_j\}$.

LMS- and RLS- GARCH(1,1) (I/V)

- LLF is maximized when predicted $\{\hat{\sigma}_t^2\}$ equal $\{r_t^2\}$

Departament of Signal Theory and Communications

LMS- and RLS- GARCH(1,1) (I/V)

Departament of Signal Theory and Communications

- LLF is maximized when predicted $\{\hat{\sigma}_t^2\}$ equal $\{r_t^2\}$
- Does this mean that the desired value for each $\hat{\sigma}_t^2$ is r_t^2?

LMS- and RLS- GARCH(1,1) (I/V)

- LLF is maximized when predicted $\{\hat{\sigma}_t^2\}$ equal $\{r_t^2\}$
- Does this mean that the desired value for each $\hat{\sigma}_t^2$ is r_t^2?
- No. Though r_t^2 is indeed an unbiased estimator of σ_t^2, we would be estimating the variance of r_t from a single realization, so r_t^2 includes a lot of noise.

LMS- and RLS- GARCH(1,1) (I/V)

- LLF is maximized when predicted $\{\hat{\sigma}_t^2\}$ equal $\{r_t^2\}$
- Does this mean that the desired value for each $\hat{\sigma}_t^2$ is r_t^2?
- No. Though r_t^2 is indeed an unbiased estimator of σ_t^2, we would be estimating the variance of r_t from a single realization, so r_t^2 includes a lot of noise.
- For GARCH(1,1) this is ok, as we are using T of these noisy estimators to fit a three-parameter model. Noise averages out.

LMS- and RLS- GARCH(1,1) (II/V)

- Can we get a better approximation to the true value of σ_t?

LMS- and RLS- GARCH(1,1) (II/V)

- Can we get a better approximation to the true value of σ_t?
- Yes, if we have access to intraday data. Sampling the closing price m times a day we can obtain the intraday return series:

$$r_{(m),t} = log\left(\frac{c_{(m),t}}{c_{(m),t-\frac{1}{m}}}\right) \quad , \quad t = \frac{1}{m}, \frac{2}{m} \dots T$$

LMS- and RLS- GARCH(1,1) (II/V)

- Can we get a better approximation to the true value of σ_t?
- Yes, if we have access to intraday data. Sampling the closing price m times a day we can obtain the intraday return series:

$$r_{(m),t} = log\left(\frac{c_{(m),t}}{c_{(m),t-\frac{1}{m}}}\right) \quad , \quad t = \frac{1}{m}, \frac{2}{m} \ldots T$$

- From $r_{(m),t}$ we can define the Cumulative Squared Returns, which are a much better estimate of σ_t^2 (but not a prediction!)

$$\text{CSR}_t = \sum_{i=0}^{m-1} r^2_{(m),t-\frac{i}{m}}$$

LMS- and RLS- GARCH(1,1) (II/V)

- Can we get a better approximation to the true value of σ_t?
- Yes, if we have access to intraday data. Sampling the closing price m times a day we can obtain the intraday return series:

$$r_{(m),t} = log\left(\frac{c_{(m),t}}{c_{(m),t-\frac{1}{m}}}\right) \quad , \quad t = \frac{1}{m}, \frac{2}{m} \dots T$$

- From $r_{(m),t}$ we can define the Cumulative Squared Returns, which are a much better estimate of σ_t^2 (but not a prediction!)

$$\text{CSR}_t = \sum_{i=0}^{m-1} r^2_{(m),t-\frac{i}{m}}$$

- GARCH predictions are indeed closer to CSR_t than to r_t^2

LMS- and RLS- GARCH(1,1) (III/V)

- We are now ready to derive an LMS-GARCH(1,1) algorithm. We can express conventional GARCH(1,1) as:

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2$$

LMS- and RLS- GARCH(1,1) (III/V)

- We are now ready to derive an LMS-GARCH(1,1) algorithm. We can express conventional GARCH(1,1) as:

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2$$

- Which can be rearranged as

$$\hat{\sigma}_t^2 = \mathbf{w}^T \begin{bmatrix} 1 \\ r_{t-1}^2 \\ \hat{\sigma}_{t-1}^2 \end{bmatrix}$$

where $\mathbf{w} = [\alpha_0 \ \alpha_1 \ \beta_1]^T$. This clarifies that we are making a recursive filtering of past data.

Departament of Signal Theory and Communications

LMS- and RLS- GARCH(1,1) (IV/V)

Departament of Signal Theory and Communications

- If we express the error at each step t as:

$$e_t = \hat{\sigma}_t^2 - \text{CSR}_t$$

we can dynamically update the weight vector $\mathbf{w}$ by using either LMS or RLS online rules.

LMS- and RLS- GARCH(1,1) (IV/V)

Departament of Signal Theory and Communications

- If we express the error at each step t as:

$$e_t = \hat{\sigma}_t^2 - \text{CSR}_t$$

  we can dynamically update the weight vector $\mathbf{w}$ by using either LMS or RLS online rules.

- This is in contrast with traditional GARCH because
  - We are minimizing MSE, instead of maximizing LLF
  - Thus, we make use of a desired output, CSR_t
  - Prediction is dynamical, instead of having alterning blocks of estimation and prediction

LMS- and RLS- GARCH(1,1) (V/V)

Some considerations about the proposed approximation:

- In RLS-GARCH(1,1), some stability issues appear if its exponential weighting parameter λ is not big enough
- Related with the above, both LMS- and RLS- GARCH(1,1) may need bounding applied to its β_1 parameter to avoid instability
- An apparent shortcoming is the need for extra data, as “high frequency” sampling is required. This should not be a problem, as this data is readily available, but a modified version of the algorithm which does not use intraday data is also tested (r_t^2 is used instead of CSR$_t$)

Experiments (I/VI)

- To evaluate the performance of the proposed algorithms versus the standard GARCH(1,1), we will make a quarterly study of the volatility of two european futures indices: IBEX35 PLUS and DJ EURO STOXX50.

Experiments (I/VI)

- To evaluate the performance of the proposed algorithms versus the standard GARCH(1,1), we will make a quarterly study of the volatility of two european futures indices: IBEX35 PLUS and DJ EURO STOXX50.
- For testing purposes we will take the eight quarters comprised in 2002 and 2003, showing both bearish and bullish markets

Departament of Signal Theory and Communications

Experiments (II/VI)

- For each quarter, we will make daily volatility predictions, and estimate the NMSE and LLF during that period.

$$\text{NMSE} = \frac{\sum_t (\hat{\sigma}_t^2 - \text{CSR}_t)^2}{\sum_t (\overline{\text{CSR}_t} - \text{CSR}_t)^2}$$

$$\text{with } \overline{\text{CSR}_t} = \frac{1}{T} \sum_t \text{CSR}_t$$

$$\text{LLF} = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left(\log \hat{\sigma}_t^2 + \frac{r_t^2}{\hat{\sigma}_t^2} \right)$$

Experiments (III/VI)

- GARCH(1,1) needs a period for training, in order to obtain an estimation of model parameters that will be applied in the test quarter.
- As training set we have used the previous quarter (3 months), the three previous quarters (9 months) and the seven previous quarters (21 months)
- This is because there is a trade off in the training length of the GARCH model
  - Longer training sets let the model learn appropiate stable-along-time parameters
  - Shorter training sets give the model the opportunity to adapt to recent changes

Experiments (IV/VI)

Department of Signal Theory and Communications

Adaptive algorithms are named as follows:

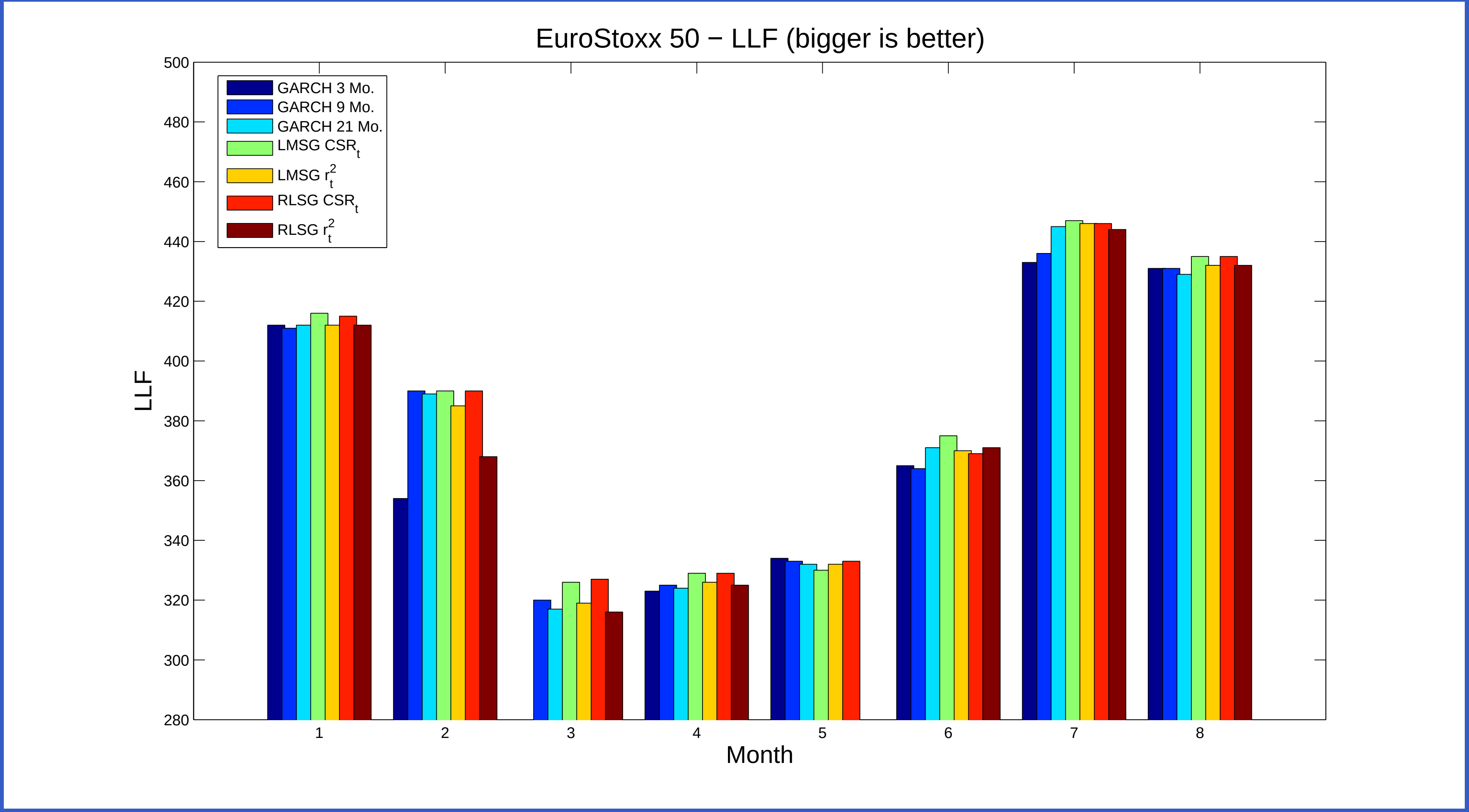
- *LMSG1* refers to the LMS-GARCH, implemented as an LMS adaptive filter, trained with MSE cost function and using CSR_t as desired values (we sample the return series every 30 min).
- *LMSG2* is the same as above but using r_t^2 as desired values.
- *RLSG1* refers to the RLS-GARCH trained with the usual RLS implementation, minimizing a weighted least squared cost function and using CSR_t as desired values.
- *RLSG2* is the same as above but using r_t^2 as desired values.

Experiments (V/VI)

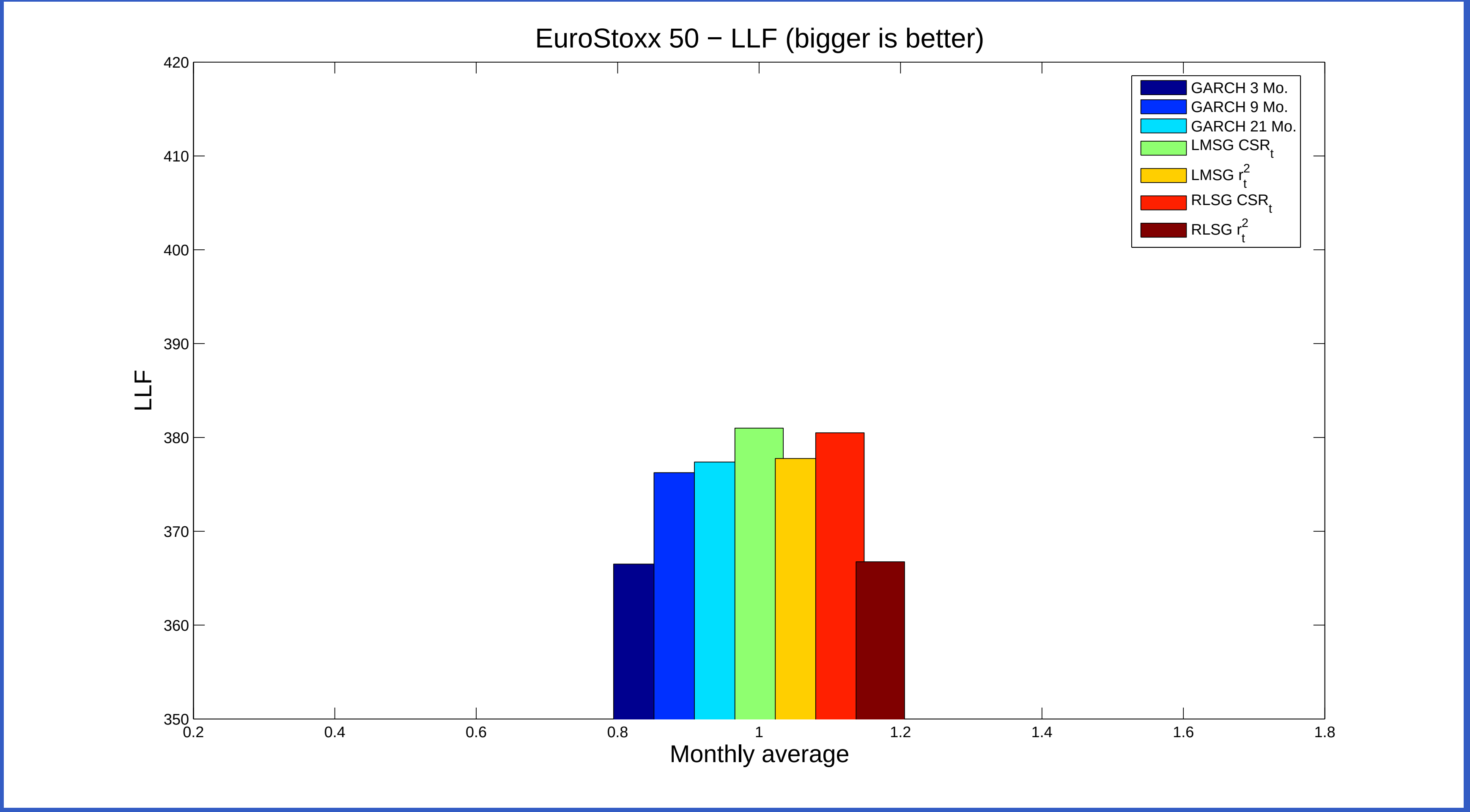
- These adaptive algorihtms need one parameter. In the LMS case, the step size μ. In the RLS case, the exponential weighting factor λ
- These parameter are fixed to suitable values using previous periods of data. Selected values are:
  - *LMSG1*, $\mu = 0.2$ and *LMSG2*, $\mu = 0.05$. In this case the selected learning step is slower for the *LMSG2* because the desired values for this algorithm are r_t^2, which are noisier than CSR$_t$.
  - *RLSG1*, $\lambda = 0.92$ and *RLSG2*, $\lambda = 0.97$. Again, more memory, or equivalently, less adaptation pace, is required for the algorithm with noisier objective

Experiments - Results

Departament of Signal Theory and Communications

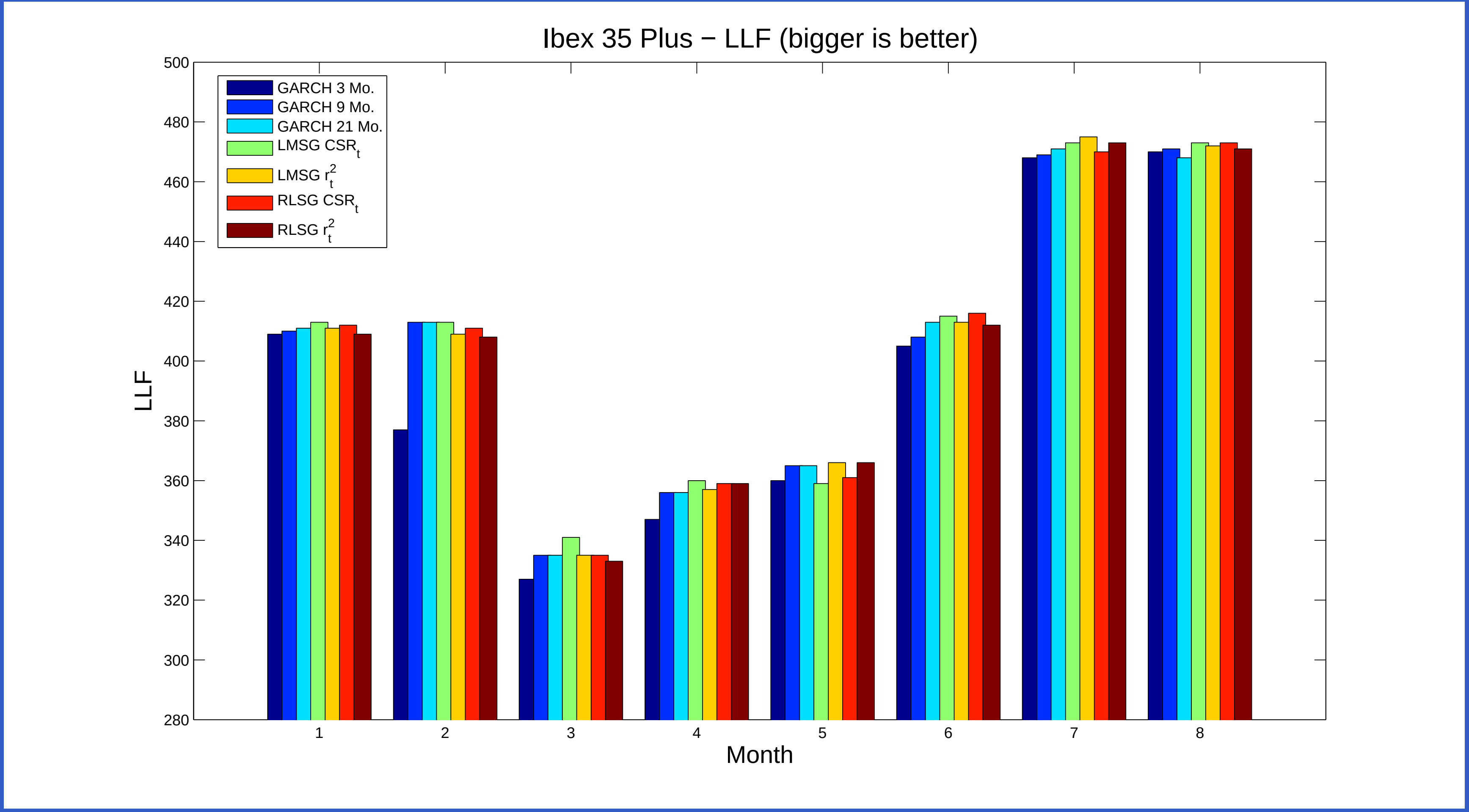


Experiments - Results

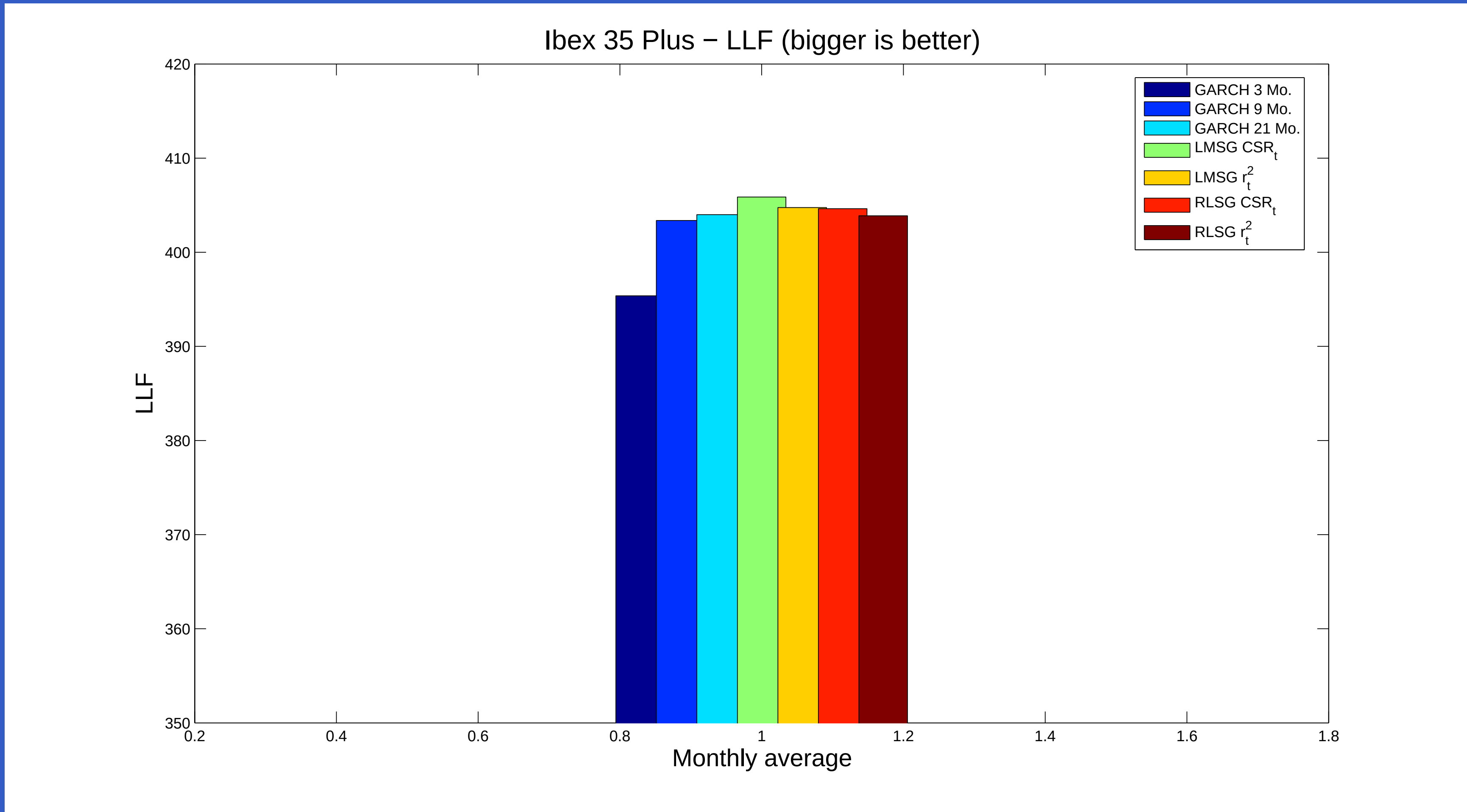


Departament of Signal Theory and Communications

Experiments - Results



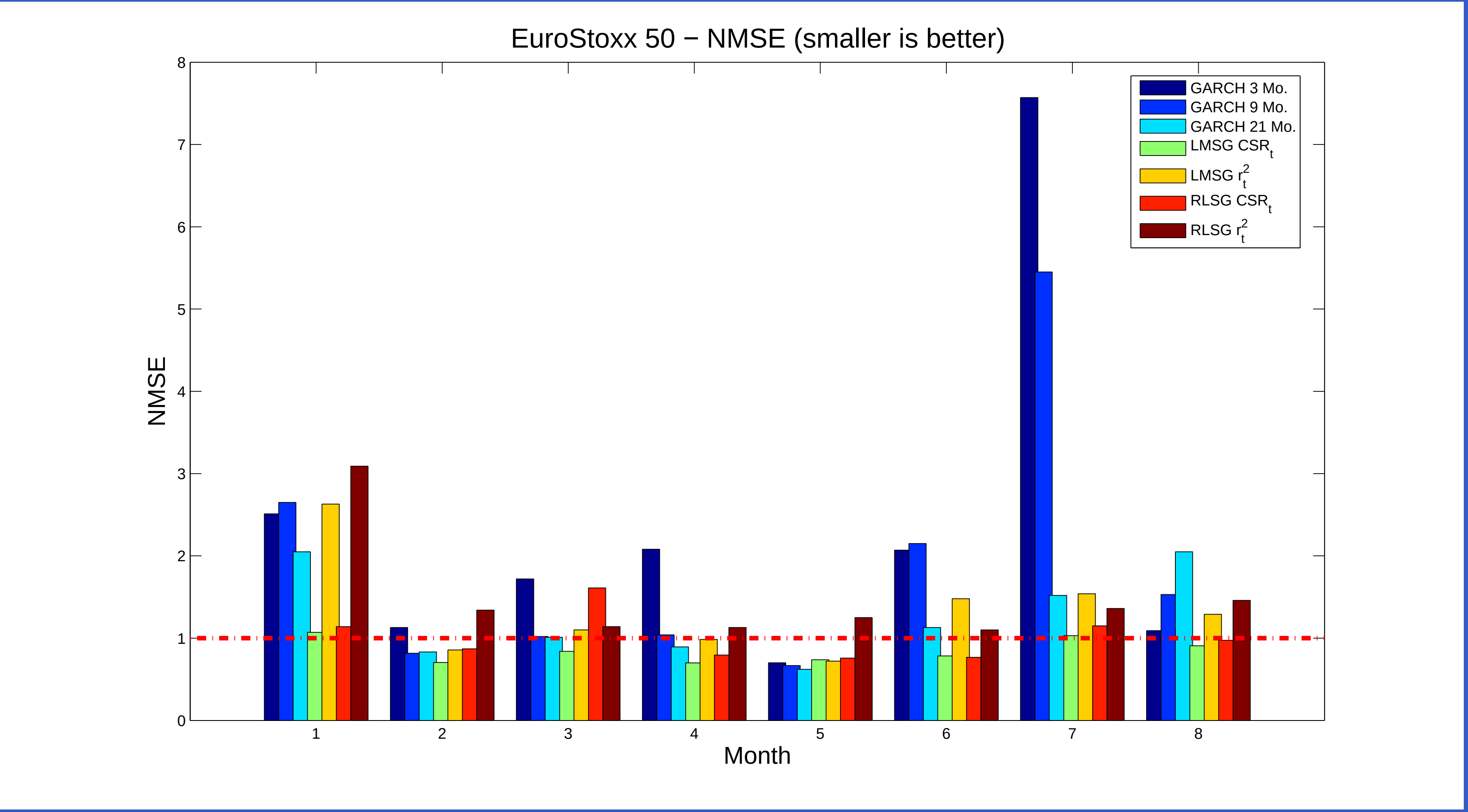
Experiments - Results



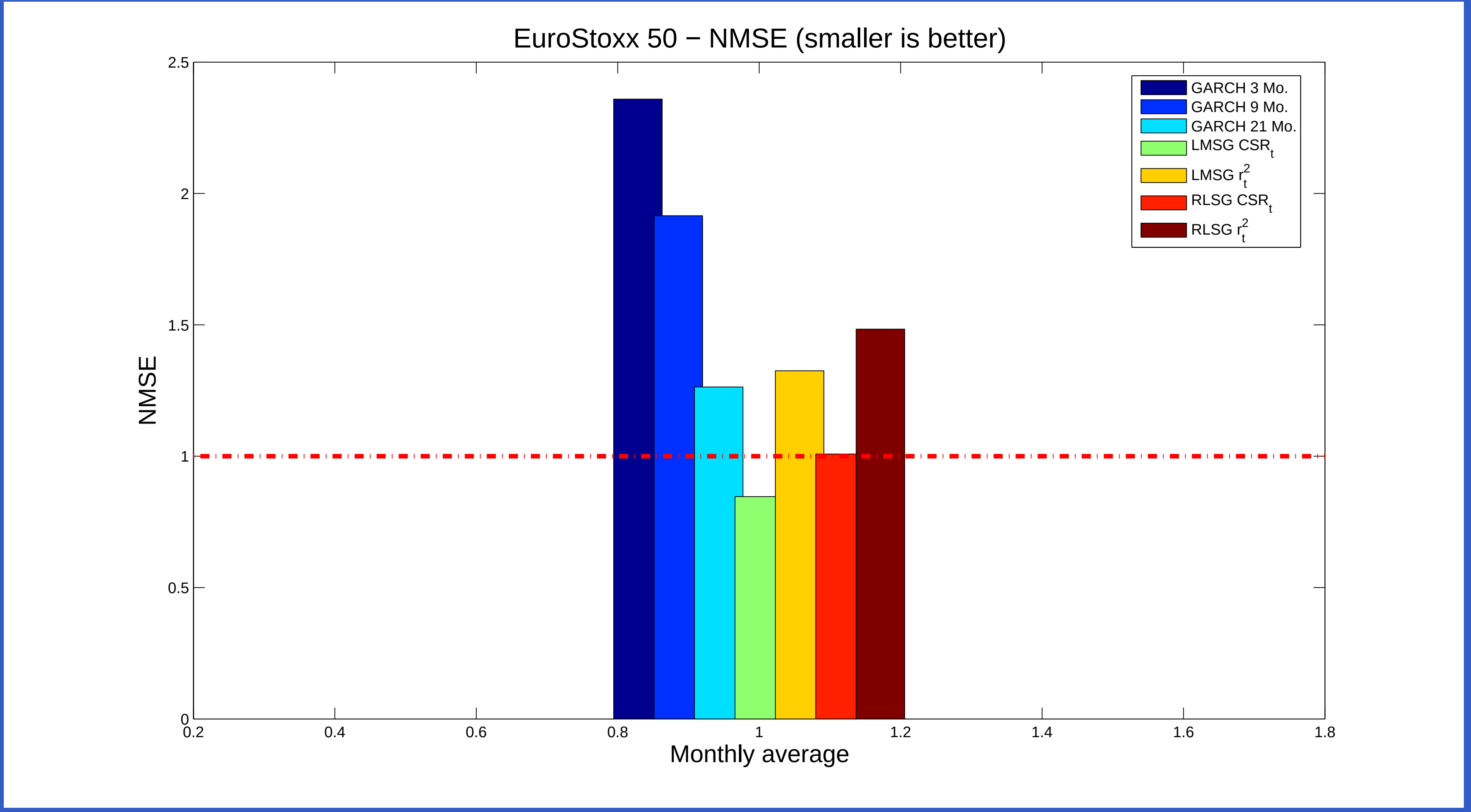
Departament of Signal Theory and Communications

Experiments - Results

Departament of Signal Theory
and Communications

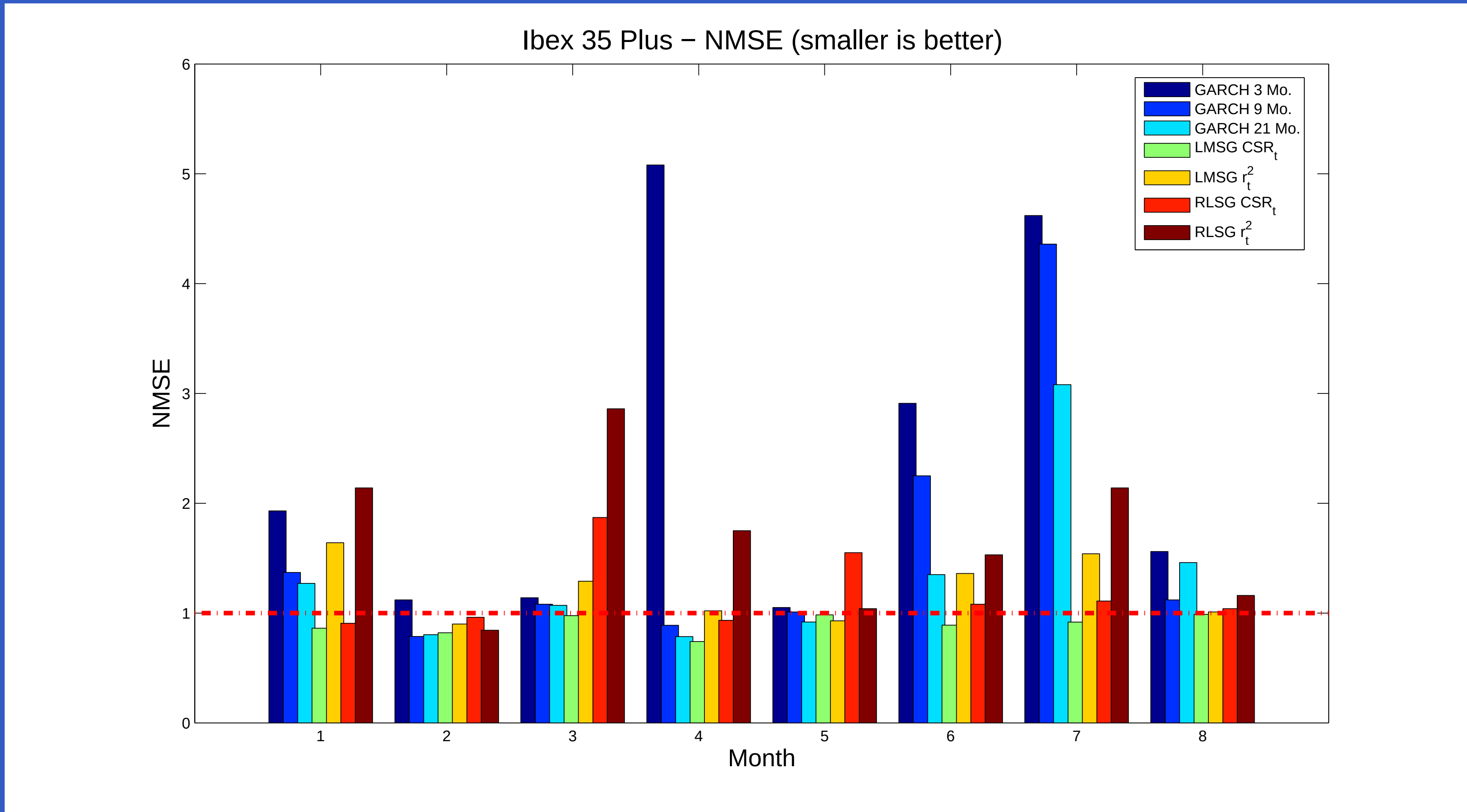


Experiments - Results



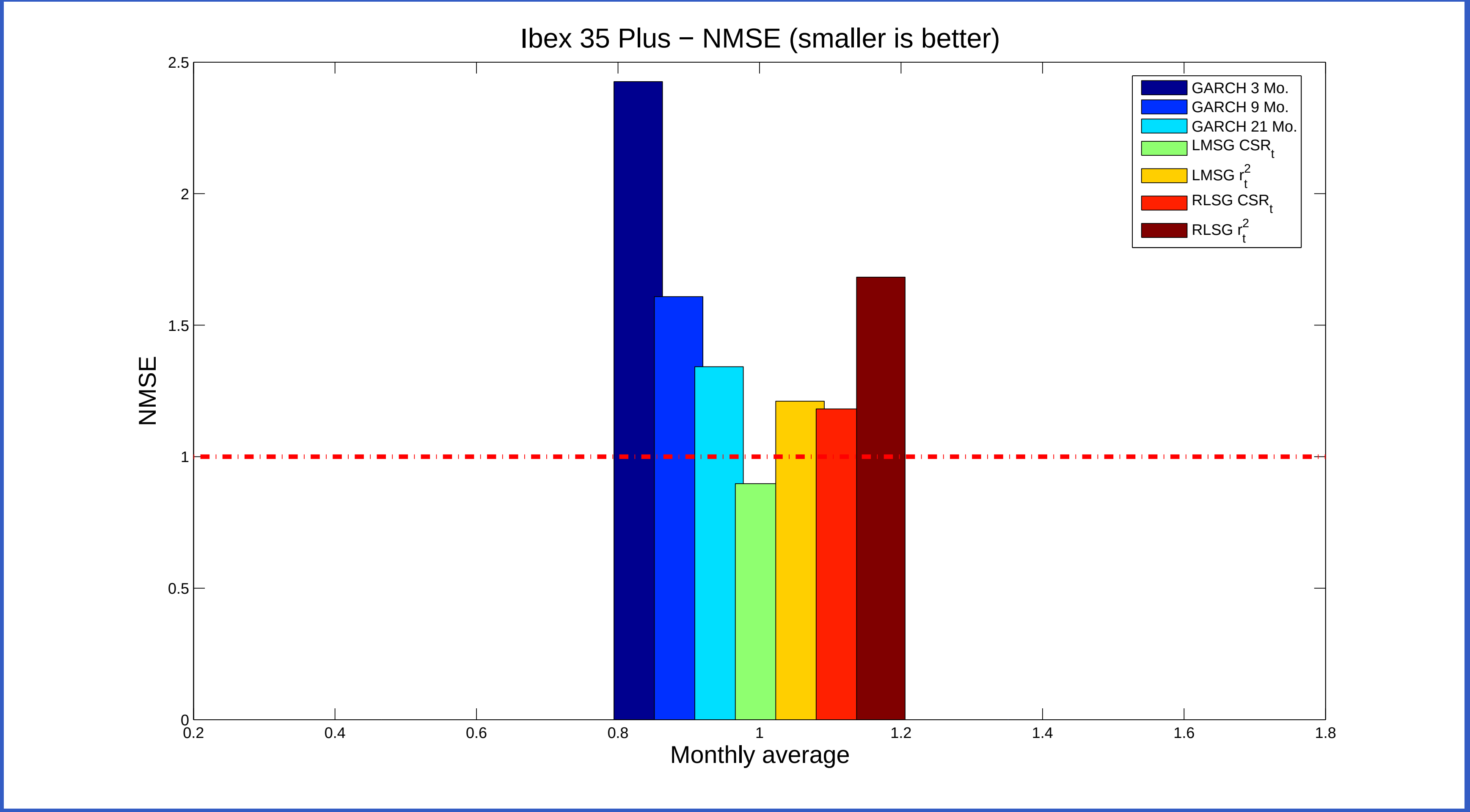
Departament of Signal Theory and Communications

Experiments - Results



Departament of Signal Theory and Communications

Experiments - Results



Experiments (VI/VI)

From the previous slides we can see that

- *LMSG1* outperforms all the other approaches.

Departament of Signal Theory
and Communications

Experiments (VI/VI)

Departament of Signal Theory and Communications

From the previous slides we can see that

- *LMSG1* outperforms all the other approaches.
- GARCH yields an NMSE above 1.0 much often than *LMSG1* does.

Experiments (VI/VI)

From the previous slides we can see that

- *LMSG1* outperforms all the other approaches.
- GARCH yields an NMSE above 1.0 much often than *LMSG1* does.
- Although new algorithms are minimizing a NMSE measure, they provide better performance in terms of LLF than GARCH, which is based on LLF.

Experiments (VI/VI)

Departament of Signal Theory and Communications

From the previous slides we can see that

- *LMSG1* outperforms all the other approaches.
- GARCH yields an NMSE above 1.0 much often than *LMSG1* does.
- Although new algorithms are minimizing a NMSE measure, they provide better performance in terms of LLF than GARCH, which is based on LLF.
- The validity of CSR$_t$ as a better σ_t estimation is validated experimentally through the results, as it provides a significant increase in accuracy for the LLF measure. (It also increases accuracy for the NMSE measure, but this could be interpreted as due to NMSE's own definition).

Conclusions

Departament of Signal Theory and Communications

GARCH(1,1) models have proved to provide accurate forecasts, and are difficult to beat by more sophisticated models. The proposed models (in particular, *LMSG1*) have the following advantages:

- Coefficients calculations are simpler. Just one LMS step per prediction is required.

Conclusions

GARCH(1,1) models have proved to provide accurate forecasts, and are difficult to beat by more sophisticated models. The proposed models (in particular, *LMSG1*) have the following advantages:

- Coefficients calculations are simpler. Just one LMS step per prediction is required.
- On-line and adaptative. There is no need to choose the length of the previous training period and, more importantly, to decide when the model is no longer valid.

Conclusions

Departament of Signal Theory and Communications

GARCH(1,1) models have proved to provide accurate forecasts, and are difficult to beat by more sophisticated models. The proposed models (in particular, *LMSG1*) have the following advantages:

- Coefficients calculations are simpler. Just one LMS step per prediction is required.
- On-line and adaptative. There is no need to choose the length of the previous training period and, more importantly, to decide when the model is no longer valid.
- Better overall accuracy.