European Geosciences Union
General Assembly 2009

Vienna, Austria,
19 - 24 April 2009

Session: CL54/NP4.5
Climate time series analysis:
Novel tools and their application

# The trendy trends: a fashion or a science story?

S.M. Papalexiou and N. Zarkadoulas

Department of Water Resources and Environmental Engineering,
National Technical University of Athens, Greece

# 1. Abstract

The Nobelist physicist Niels Bohr once said that prediction is very difficult, especially if it is about the future. Nowadays, the scene has changed. It seems that since the scientific community accepted, in its majority, that the earth's climate is rapidly changing, an opinion that also echoes in public, scientists all over the world have identified significant trends in many climate related processes, e.g., global temperature, rainfall, river discharges, ice melting etc. Furthermore, if we adopt the suggested trends in those natural processes and their future projections we should expect a horrifying future. But is that so? How consistent and scientifically sound are these trend based scenarios? A trend in its most common form can be expressed as a linear regression line fitted to an observed sample of the natural process under investigation. In addition, the decision of whether or not a trend is significant is based on inferences regarding the regression line coefficients. However, classical statistics inferences of the regression line coefficients assume normal and independent data, assumptions that are generally not valid in natural processes. Particularly, while the assumption of normality may hold in some cases, it is well documented that natural processes exhibit a great variety of autocorrelation structures, exponential or power type, and thus the assumption of independently distributed data is violated. In this study, we investigate based on Monte Carlo simulations the effect of different autocorrelation structures in the inference of the trend line significance. We demonstrate that trends considered as significant in a classical statistics framework are actually insignificant if autocorrelation structures are incorporated.

# 2. Motivation

It was 2500 year ago when the famous ancient Greek philosopher Heraclitus of Ephesus said that "πάντα ῥεῖ", i.e., everything is under constant change. Nevertheless, and while it is hard to imagine a natural process that does not vary, mildly or severely, in time, the nature's "picture" that the majority of scientists have adopted nowadays, seems to be that of a peaceful lake. Therefore, events that frequently or largely deviate from the average, human-centered, ideal "picture", suggest an alarming change of the whole scheme. This is exactly the case in our time, where the agreement in a rapidly changing climate by the scientific community seems to have triggered a trend-hysteria in all climate related processes. The table on the right verifies this.

The purpose of this study is simple, i.e., to investigate linear trends in samples generated from autocorrelated stationary processes with different autocorrelation structures, and to exhibit that the presence of autocorrelation makes the difference, that is, significant trends, assuming independence in the observed sample, are actually insignificant and very probable in autocorrelated processes.

| | Results number | |
|---|---|---|
| Keyword | Google | Google Scholar |
| rainfall trends | 1,670,000 | 408,000 |
| "rainfall trends" | 13,100 | 1,720 |
| "sea level" trends | 1,400,000 | 170,000 |
| "sea level trends" | 13,800 | 1,320 |
| "river flow" trends | 89,400 | 20,900 |
| "river flow trends" | 211 | 36 |
| "river runoff" trends | 20,600 | 6,300 |
| "river runoff trends" | 232 | 10 |
| temperature trends | 17,900,000 | 2,400,000 |
| "temperature trends" | 123,000 | 12,300 |
| hurricane trends | 3,450,000 | 41,600 |
| "hurricane trends" | 1790 | 141 |
| tornado trends | 825,000 | 19,900 |
| "tornado trends" | 454 | 24 |
| "malaria trends" | 3,120 | 843 |
| malaria trends | 1,890,000 | 96,500 |

# 3. Trends vs. stationarity

Let $\{X(t), t \in T\}$ be a stationary normal process with mean $\mu_X$ and standard deviation $\sigma_X$, and $\{x_i\}_n$ a random sample of size $n$ and $x = \alpha\ t + \beta$ a linear regression line, with slope $\alpha$ and intercept $\beta$, fitted to the random sample. As the process $\{X(t)\}$ is stationary, i.e., the joint density of any order remains the same in time, the slope $\alpha$, clearly, will tend to zero and the intercept $\beta$ will tend to the mean of the process $\mu_X$ as the sample size $n$ tends infinity. Nevertheless, for finite sample sizes and for all possible realizations of the process, the estimator of the slope $A$, and the estimator of the intercept $B$, are random variables (r.v.'s). Specifically, it is well known (e.g., Soong, 2004, ch. 11 ) that when the process $\{X(t)\}$ is normal, the r.v.'s $A$ and $B$ form a bivariate normal r.v. $(A,B)$ with joint probability density function (pdf) $f_{AB}(\alpha,\beta;\ \mu_A,\ \mu_B,\ \sigma_A,\ \sigma_B,\ \rho_{AB})$, where $\mu_A$, $\mu_B$, $\sigma_A$, $\sigma_B$ are the means and standard deviations of $A$ and $B$, respectively, and $\rho_{AB}$ is the correlation coefficient of $A$ and $B$. Clearly, the parameters of the joint pdf of the r.v. $(A,B)$, depend on the mean $\mu_X$, the standard deviation $\sigma_X$, the autocorrelation structure of $\{X(t)\}$, and the sample size $n$. Additionally, the conditional pdf of the r.v. $A$, for a given value of $B$, is given by $f_{A|B}(\alpha|\beta) = f_{AB}(\alpha,\beta)/\ f_B(\beta)$, that can be easily proven (e.g., Feller, 1970, p. 72) that is a normal pdf given by $f_{A|B}(\alpha|\beta;\ \mu_{A|B},\sigma_{A|B})$, with $\mu_{A|B}= \mu_A + \rho_{AB}(\sigma_A/\sigma_B)(\mu_B + \beta)$ and $\sigma_{A|B} = (1 - \rho_{AB})^{0.5}\ \sigma_A$. As a consequence, confidence intervals (CI) for the slope $A$, can be easily estimated using the quantile function of the afore-mentioned distribution, i.e., $Q_{A|B}(u;\ \mu_{A|B},\sigma_{A|B})$.

# 4. The Monte Carlo simulation scheme

Short-term persistence process

**AR(1) synthetic series generation**

Model statistics: $\mu_X = 0$, $\sigma_X = 1$

Number of series: 10000

*Lag*-1 autocorrelation coefficients:

$\boldsymbol{\rho}_1 = \{0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.99\}$

Sample sizes:

$\boldsymbol{n} = \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}$

Long-term persistence process

**FGN synthetic series generation**

Model statistics: $\mu_X = 0$, $\sigma_X = 1$

Number of series: 10000

Hurst exponents:

$\boldsymbol{H} = \{0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$

Sample sizes:

$\boldsymbol{n} = \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}$

Regression coefficients estimation for every synthetic series

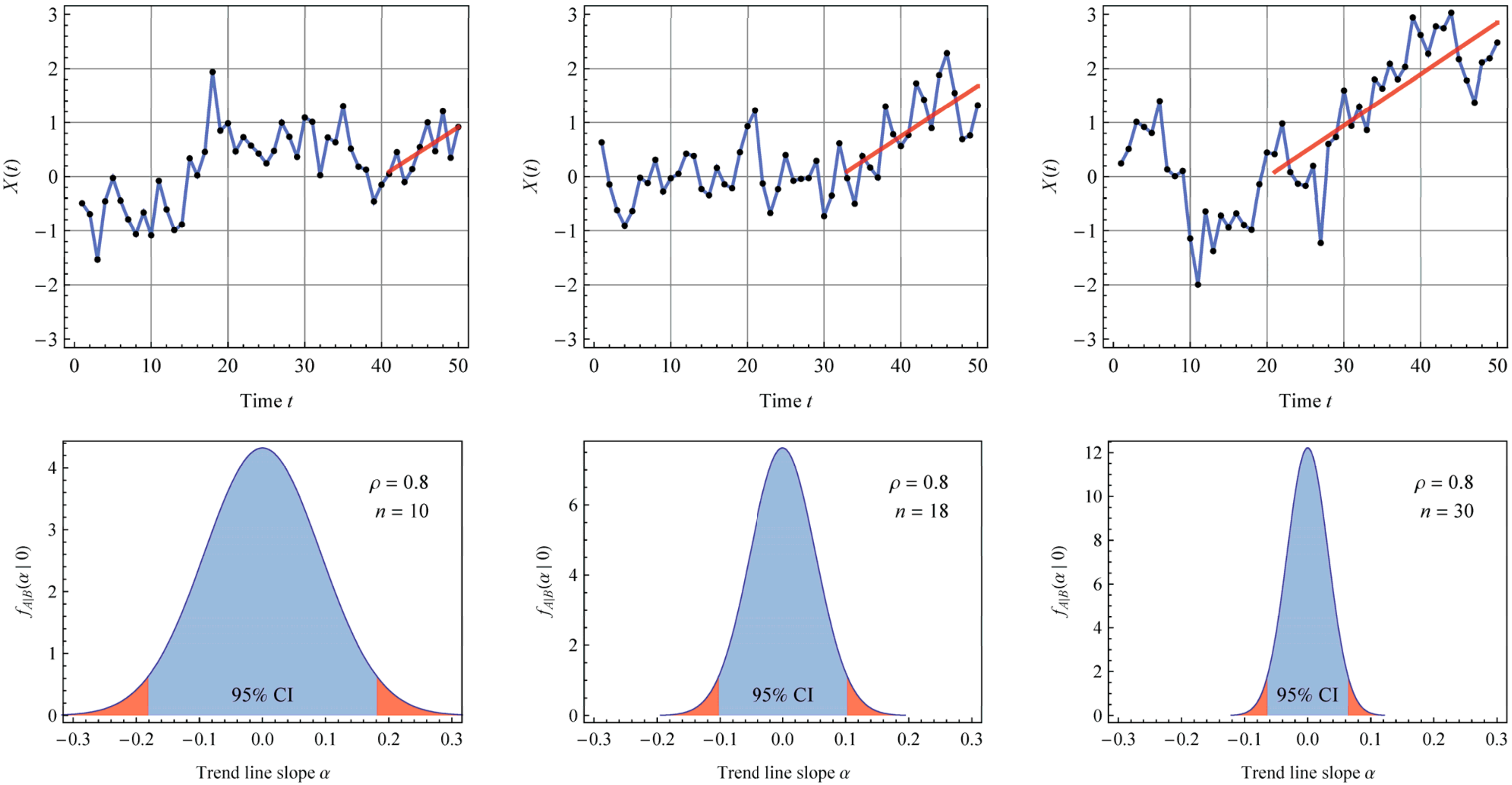
Slope $A$ and intercept $B$ statistics estimation for every pair of $(\rho_1, L)$ and of $(H, L)$

Estimated statistics: means $\mu_A$, $\mu_B$, standard deviations $\sigma_A$, $\sigma_B$, and correlation coefficient $\rho_{AB}$

Construction of interpolation functions $\sigma_A(\rho_1,n)$, $\sigma_B(\rho_1,n)$, $\rho_{AB}(\rho_1,n)$ and thus of the distribution function $F_{AB}(\alpha,\beta;\, \mu_A,\, \mu_B,\, \sigma_A,\, \sigma_B,\, \rho_{AB})$

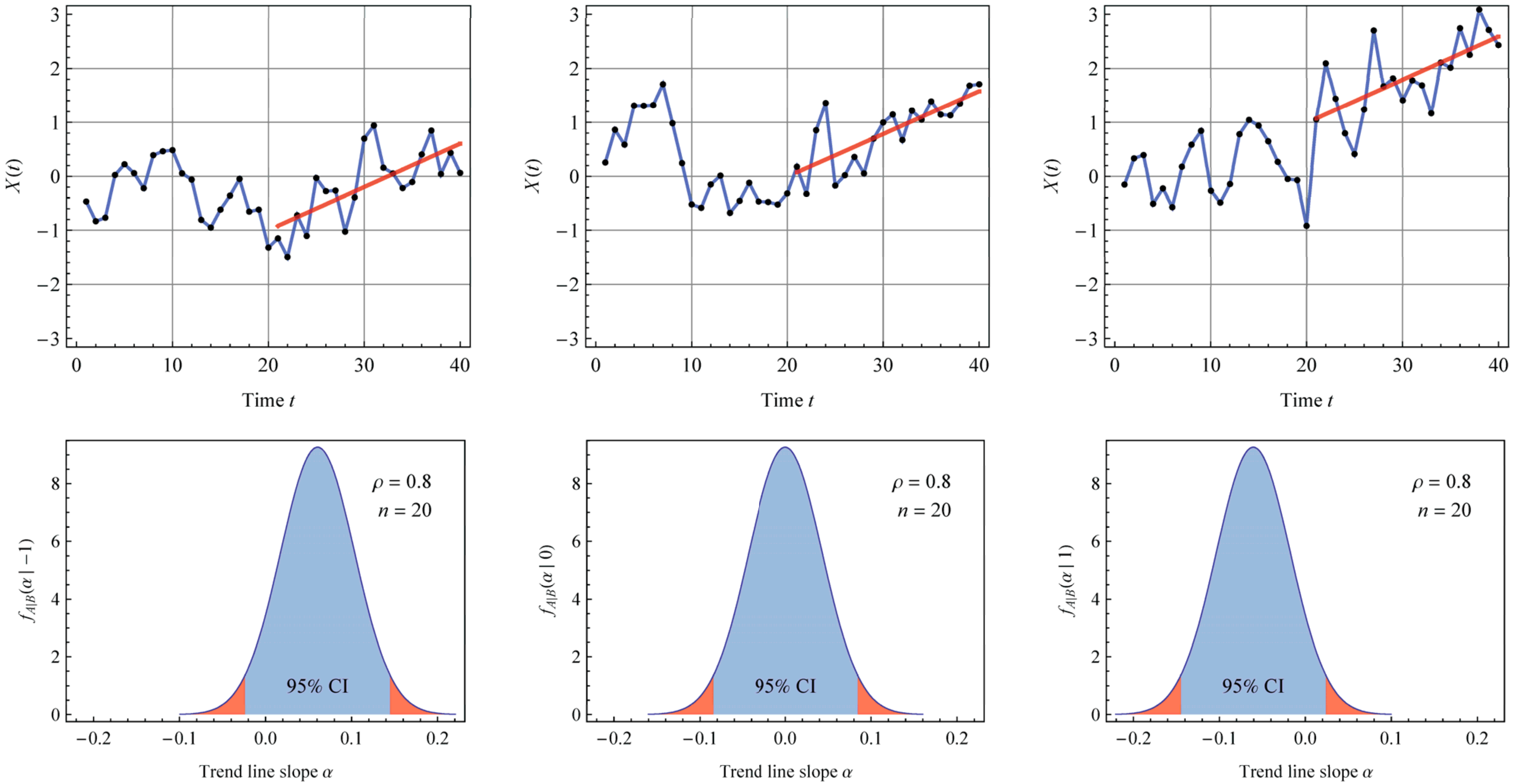
Construction of interpolation functions $\sigma_A(H,n)$, $\sigma_B(H,n)$, $\rho_{AB}(H,n)$ and thus of the distribution function $F_{AB}(\alpha,\beta;\, \mu_A,\, \mu_B,\, \sigma_A,\, \sigma_B,\, \rho_{AB})$

# 5. Slope vs. sample size



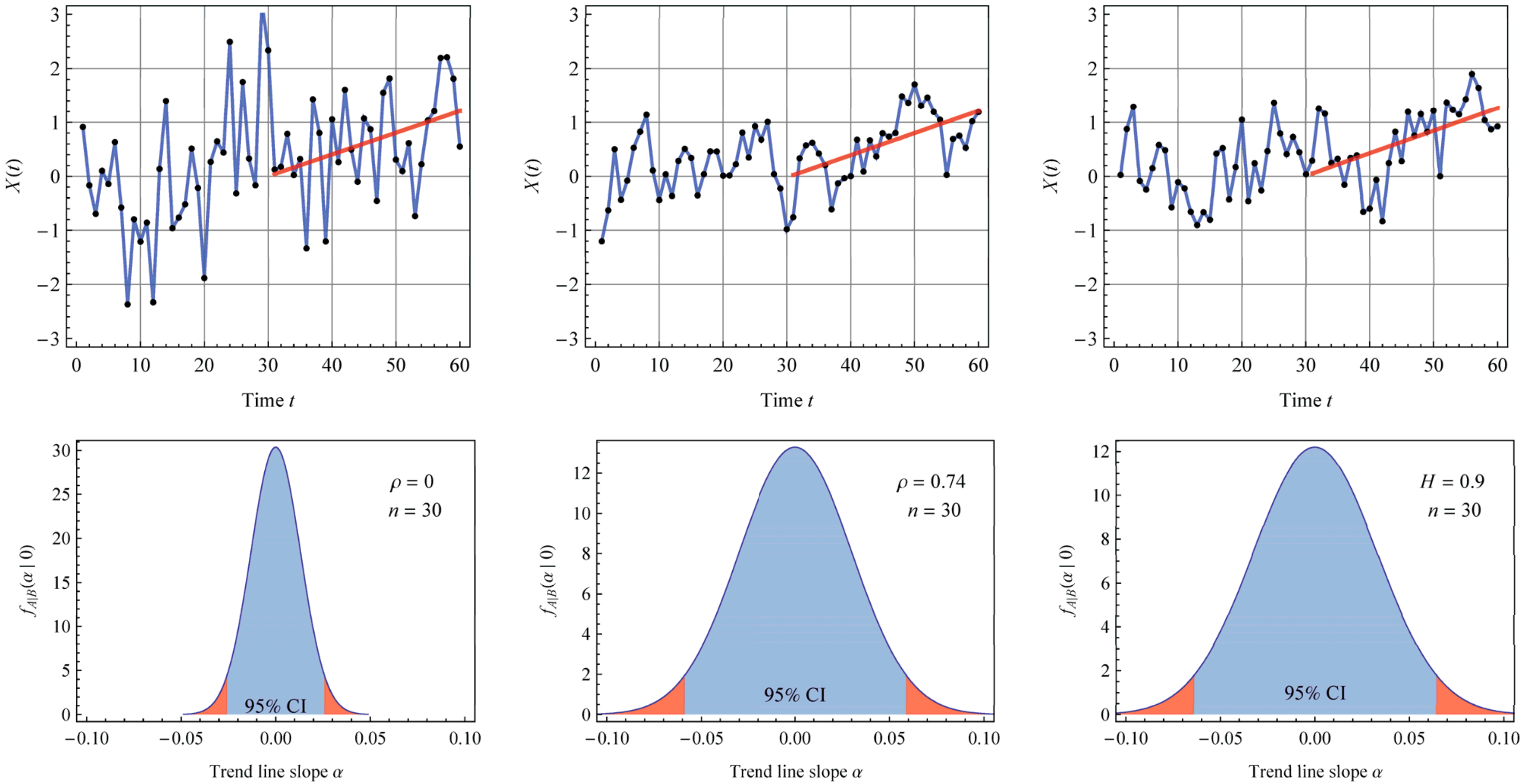

Three realizations of an AR(1) process with *lag*-1 autocorrelation coefficient $\rho_1 = 0.8$ and the fitted trend lines to the last 10, 18 and 30 values. The regression coefficients are the same in all cases, i.e., $\alpha \approx 0.1$ and $\beta \approx 0$, but as the slope's conditional distribution reveals, the exceedence probability $F^*_{A|B}(\alpha|0) = 1\text{-}\ F_{A|B}(\alpha|0)$ in the $n = 10$ case is high, in the $n = 18$ case is approximately $F^*_{A|B}(\alpha|0) \approx 2.5\%$ and in the $n = 30$ case the $F^*_{A|B}(\alpha|0)$ is very small.

# 6. Slope vs. vertical-axis intercept



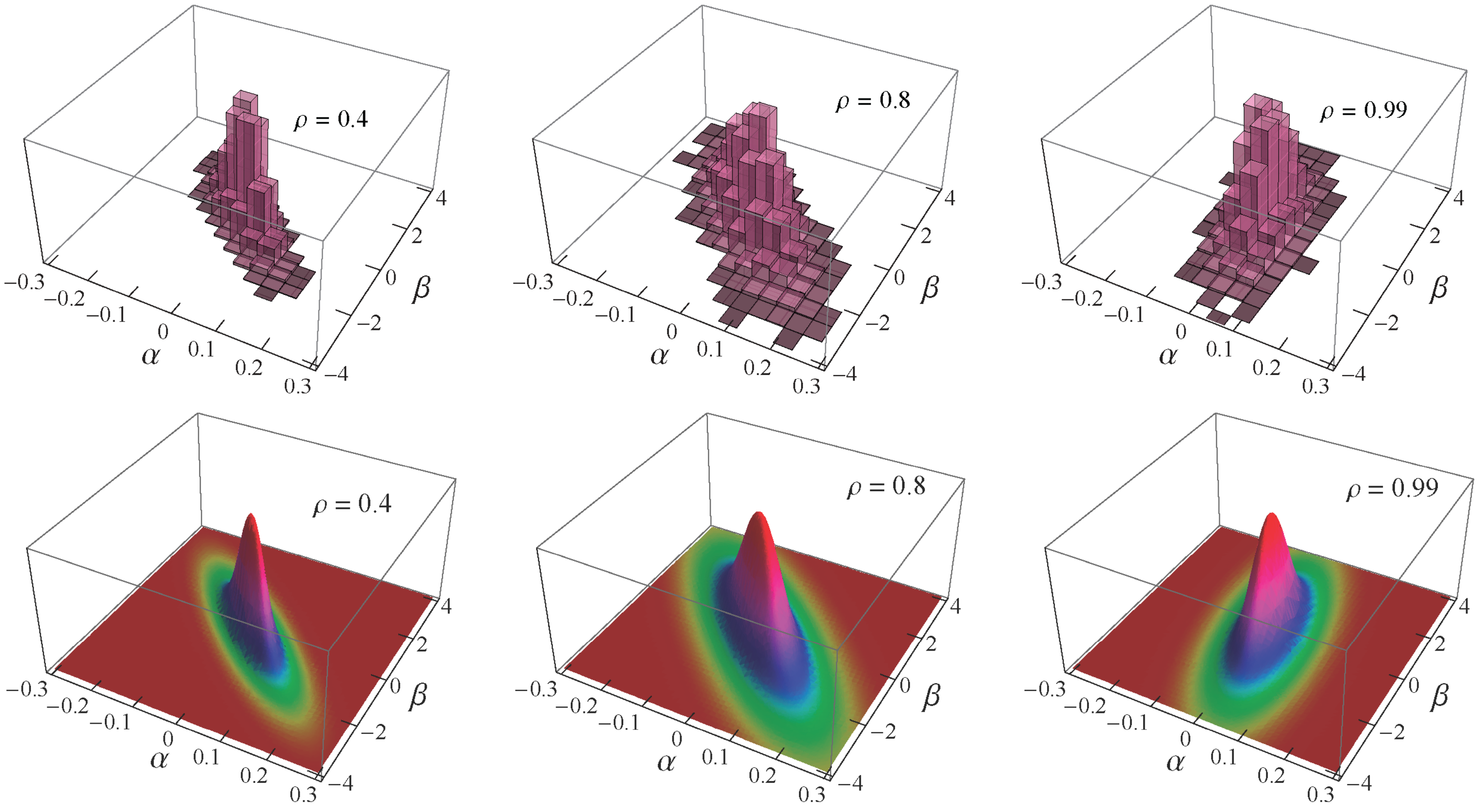

Three realizations of an AR(1) process with *lag*-1 autocorrelation coefficient $\rho_1 = 0.8$ and the fitted trend lines to the last 20 values. The slope is the same in all cases, i.e., $\alpha \approx 0.08$ and the intercepts are $\beta \approx$ -1, $\beta \approx 0$ and $\beta \approx 1$ (left to right). Inspection of the slope's conditional distribution reveals that the exceedence probability $F^*_{A|B}(\alpha|\text{-}1)$ is high, $F^*_{A|B}(\alpha|0) \approx 2.5\%$ while the $F^*_{A|B}(\alpha|1)$ is very small.

# 7. Slope vs. autocorrelation



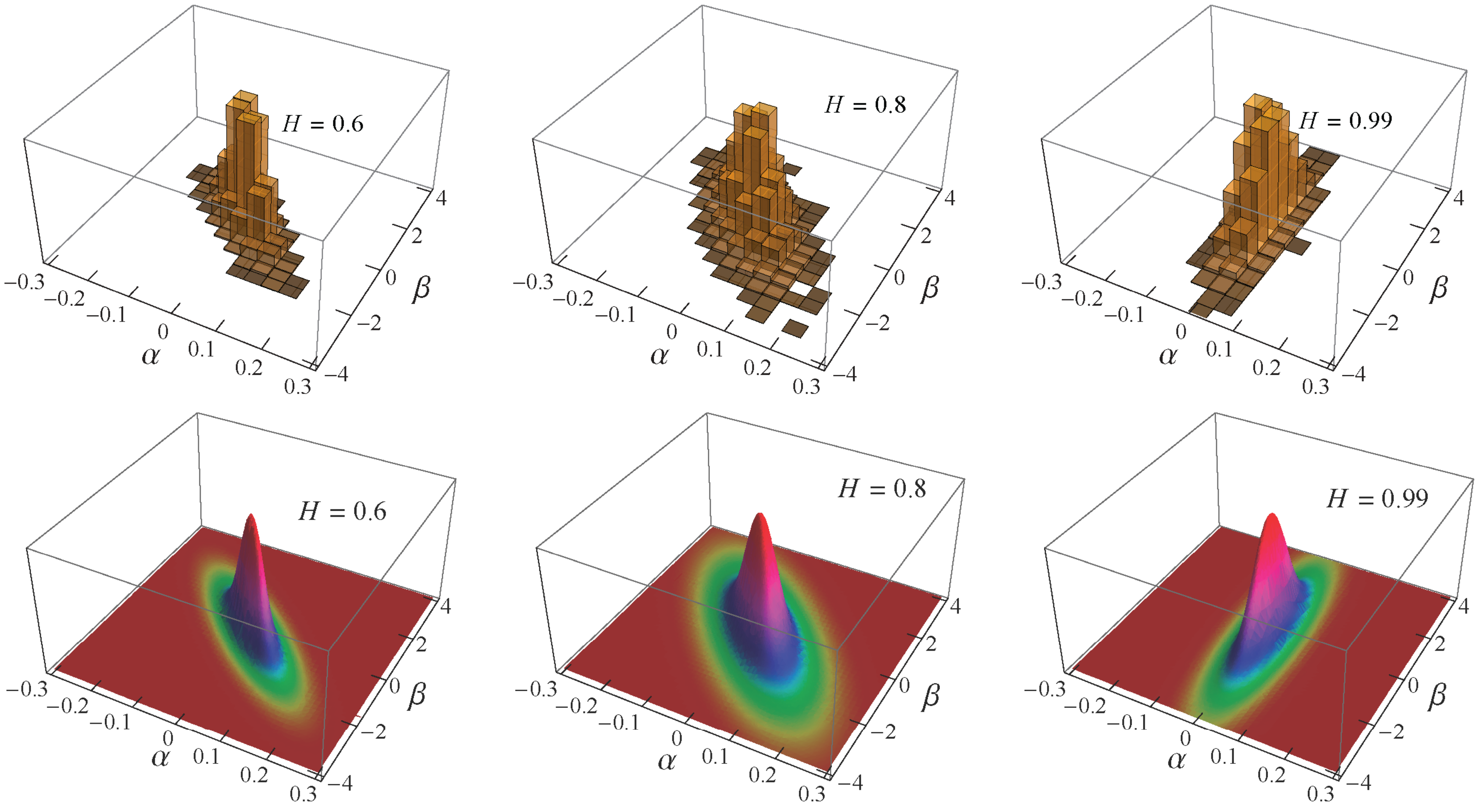

The first sample was generated from an independent process, the second from an AR(1) process with $\rho_1 = 0.74$ and the third from an FGN process with Hurst exponent $H = 0.9$ (the *lag*-1 correlation coefficient of an FGN process with $H = 0.9$ is 0.74). The regression coefficients are the same in all cases, i.e., $\alpha \approx 0.04$ and $\beta \approx 0$. The $F^*_{A|B}(\alpha\,|\,0)$ in the first case is very small while in the AR(1) and the FGN cases are very high.

# 8. AR(1): Empirical and theoretical joint densities



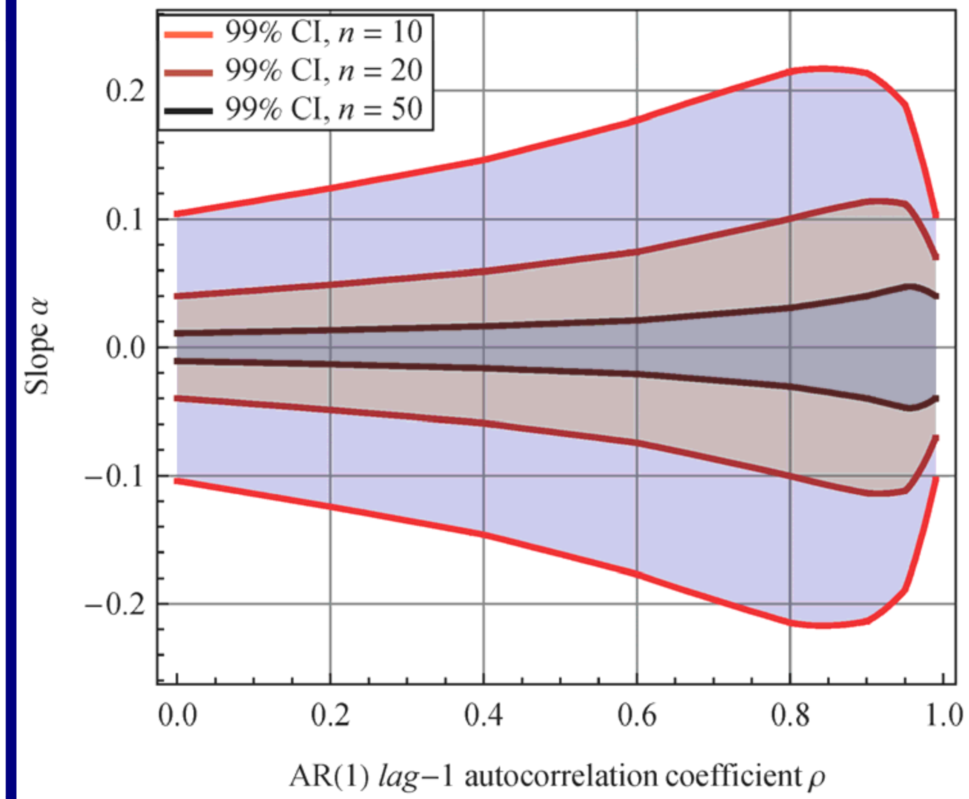

The empirical joint densities of the regression coefficients $\alpha$ and $\beta$ for three AR(1) processes with different *lag*-1 autocorrelation coefficients as resulted from the simulation of 10000 samples of size $n$ = 20 for each process, and the corresponding theoretical fitted joint densities.

# 9. FGN: Empirical and theoretical joint densities



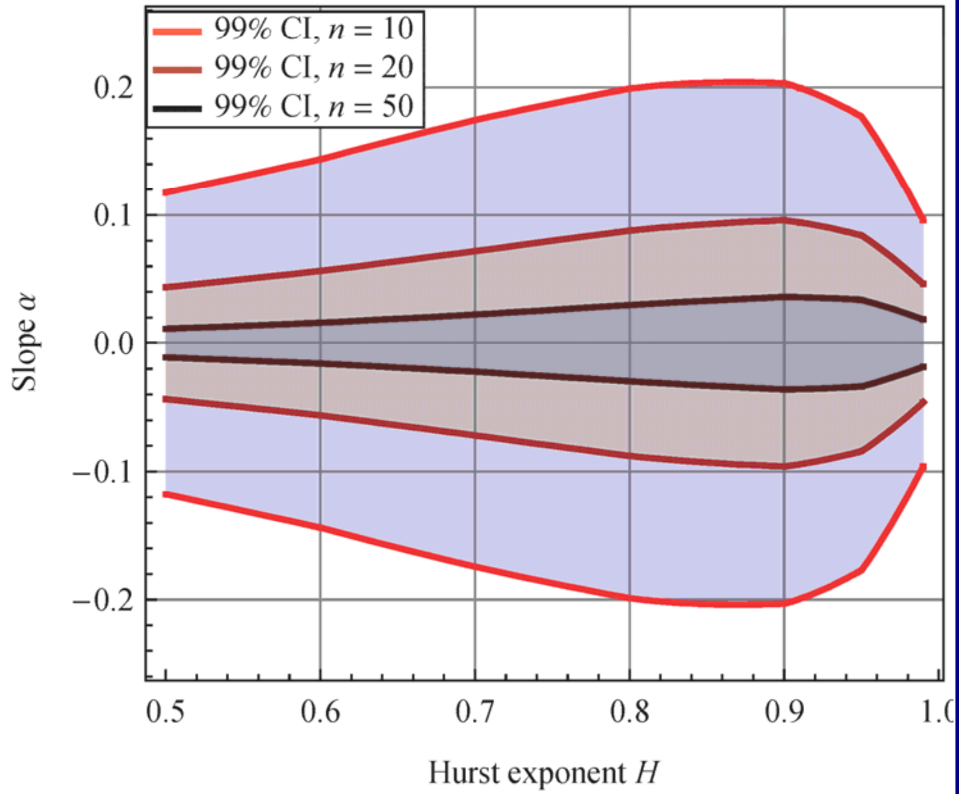

The empirical joint densities of the regression coefficients $\alpha$ and $\beta$ for three FGN processes with different values of the Hurst exponent as resulted from the simulation of 10000 samples of size $n = 20$ for each process, and the corresponding theoretical fitted joint densities.

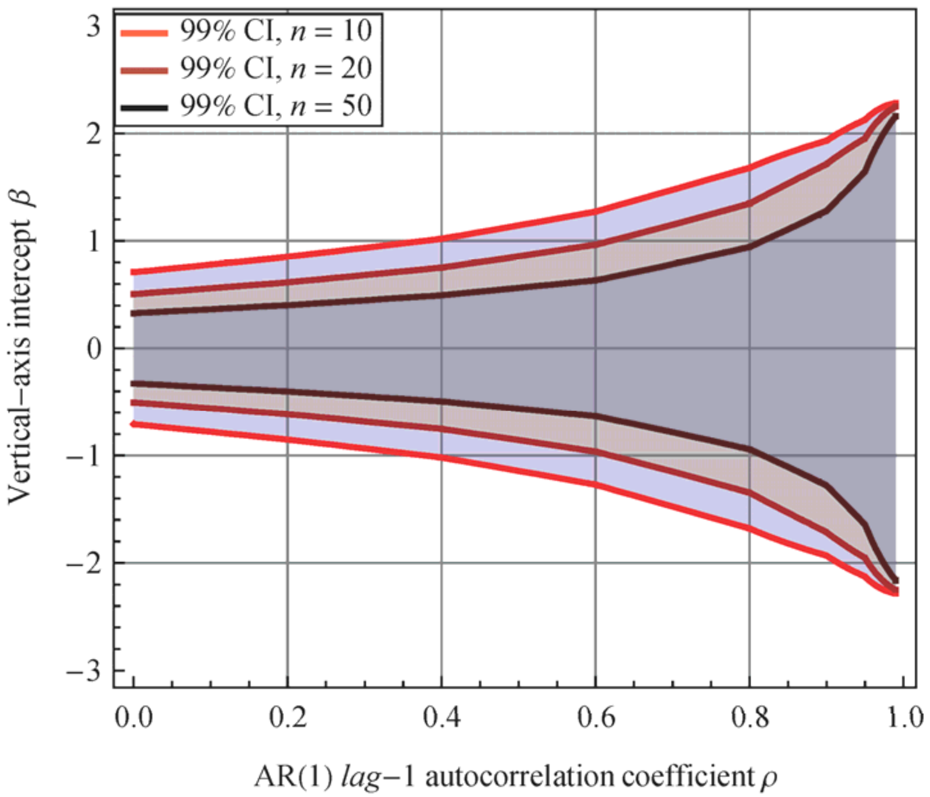
# 10. The “Claw effect” and the “Funnel effect”




## The “Claw effect”

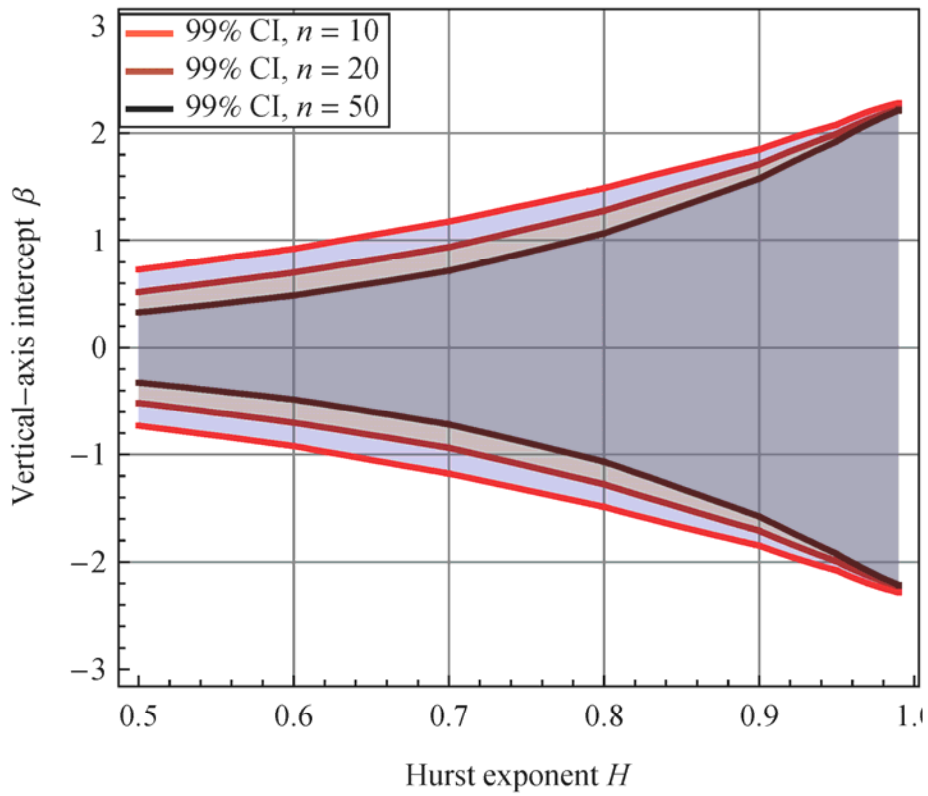
The figures depict the 99% CI of the slope $\alpha$, given that the intercept $\beta$ = 0, of an AR(1) process (left) and of an FGN process (right) vs. the autocorrelation intensity. In both processes, the CI forms a shape that resembles a claw, i.e., the CI gets wider up to a point as the intension of the autocorrelation structure increases, and then starts to narrow.



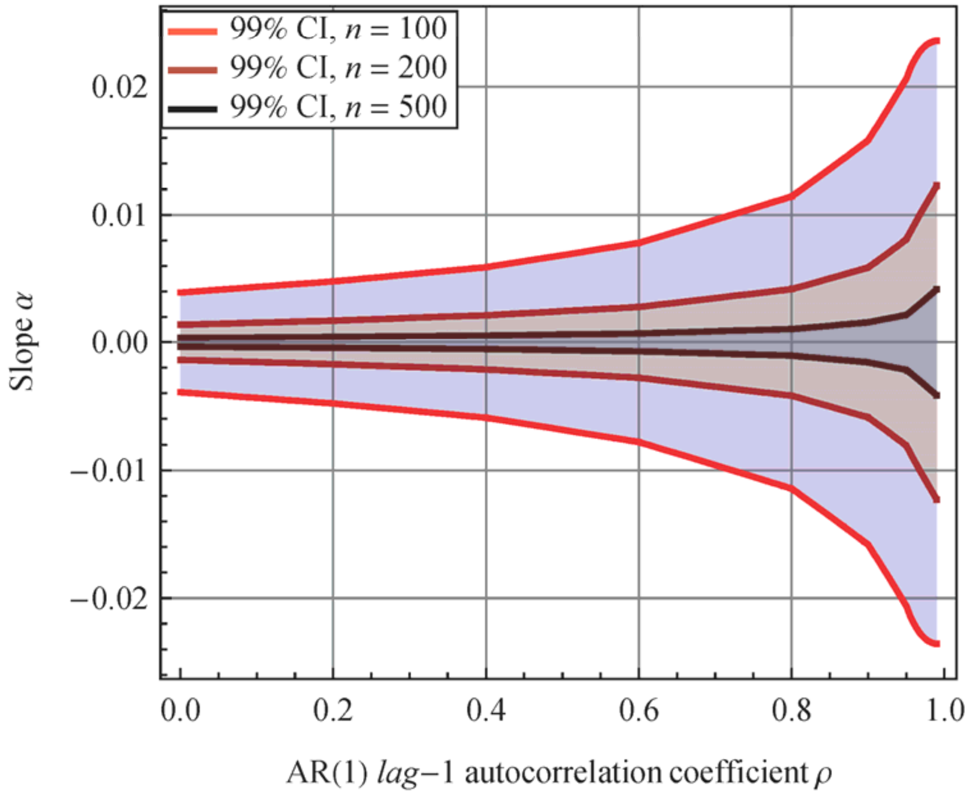




## The “Funnel effect”

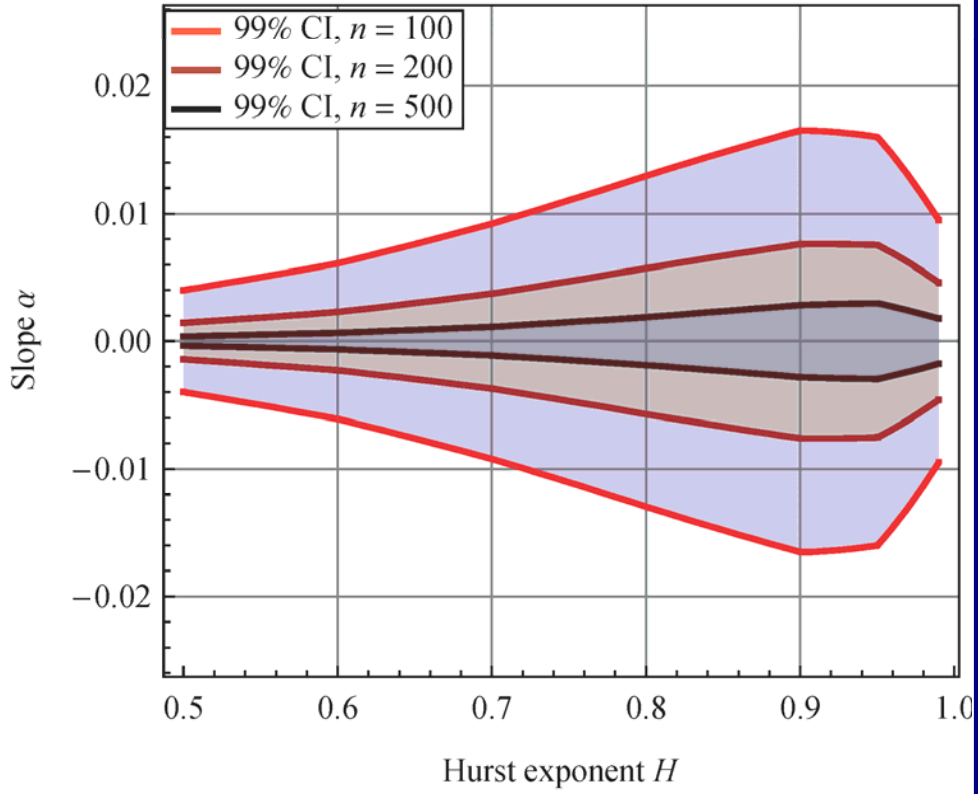
The figures depict the 99% CI of the intercept $\beta$, given that the slope $\alpha$ = 0, of an AR(1) process (left) and of an FGN process (right) vs. the autocorrelation intensity. In both processes, the CI forms a shape that resembles a funnel, i.e., the CI gets wider as the intension of the autocorrelation structure increases.

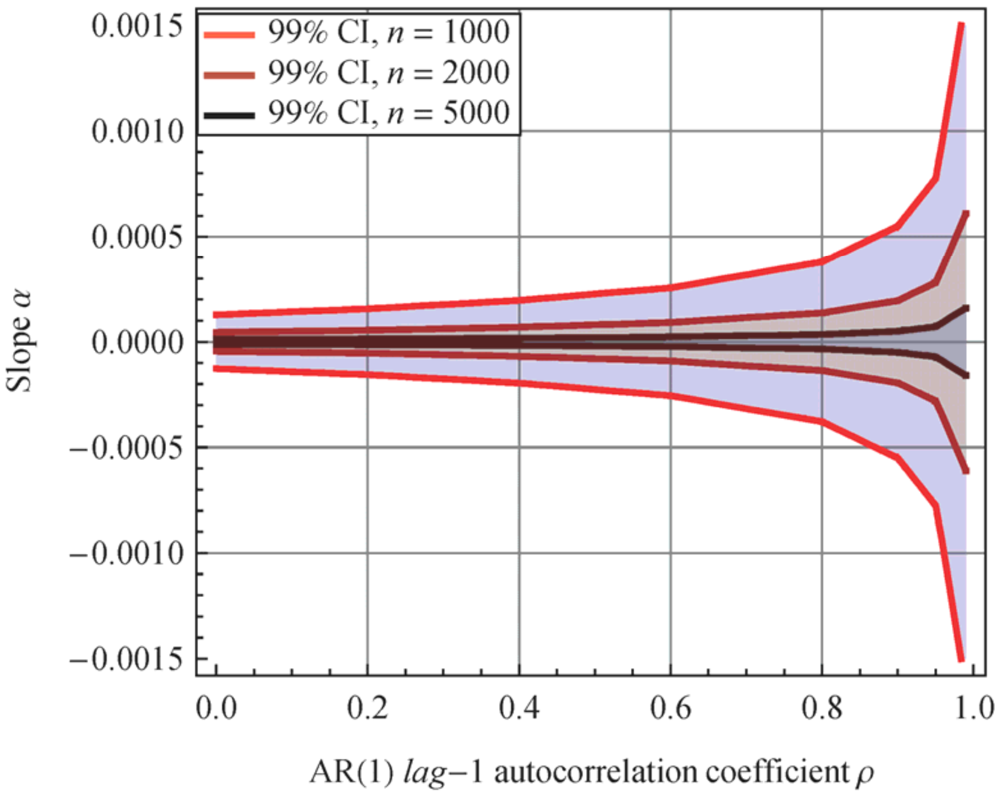


# 11. The “Claw effect” in large samples

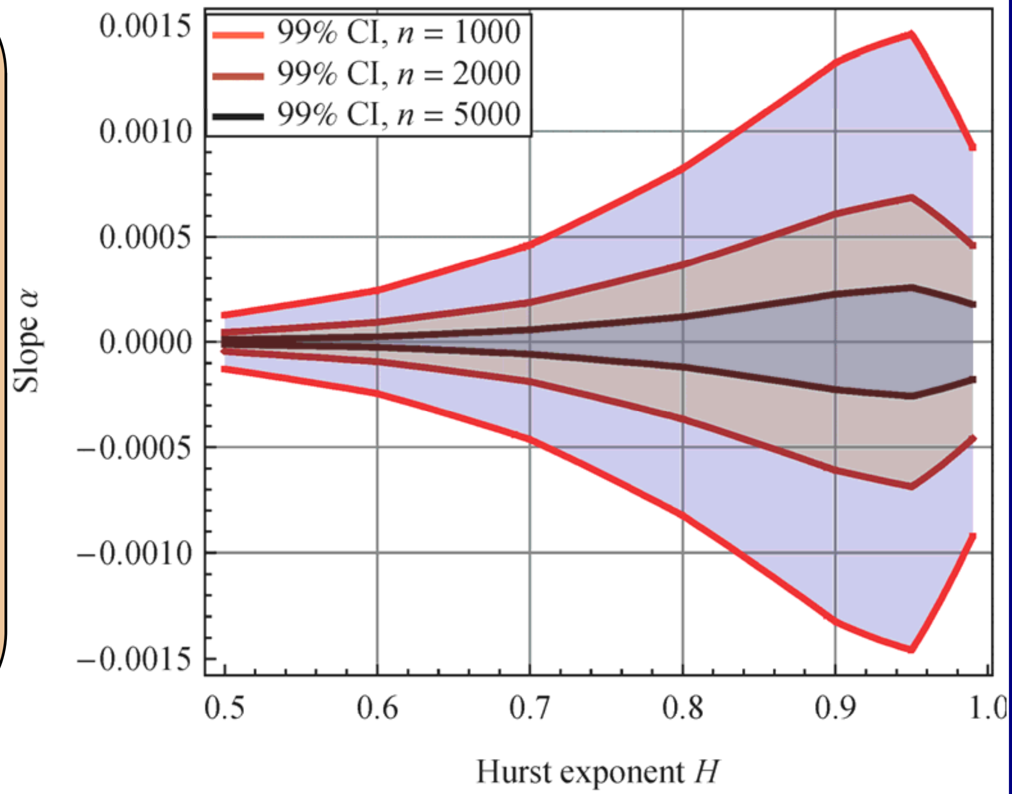



It seems that for moderate and larges sample sizes, the “Claw effect” in the slope’s CI vanishes in the AR(1) process (left) and is replaced by the “Funnel” effect. In contrast, it remains in the FGN process (right), but it seems that value of $H$ where the decrease of the CI range starts gets higher as the sample size increases.




It is worthy to note, that as the sample size increases and so the slope’s variance and the CI range decrease, the difference between the minimum CI range, that appears for $H = 0.5$ or for $\rho = 0$ (independent data), and the maximum CI range, that appears for a certain value of $H$ or $\rho$, also increases.



# 12. Conclusions

- The regression line coefficients, i.e., the slope $\alpha$ and the intercept $\beta$, for the whole set of size-$n$ realizations generated from a stationary stochastic process, form a bivariate normal distribution that its characteristics depend not only on the size $n$, but also on the autocorrelation structure of the process.
- A corollary of the previous conclusion, is that the distribution of the slope $\alpha$ is not only sample-size dependent, but also, strongly depends on the intercept $\beta$ value and on the process's autocorrelation structure.
- The effect of the autocorrelation structure in the slope variance is enormous. For certain autocorrelations structures the resulted slope variance value may be more than double compared to the variance value in the independent case. Thus, slope values with very small exceedence probability in independent data have actually high exceedence probability if the data are autocorrelated.
- This study revealed a phenomenon we named the "Claw effect", that is, the standard deviation or the $u$% CI range of the slope vs. the autocorrelation intensity, form a shape that resembles a claw, i.e., their values increase up to a point as the intension of the autocorrelation structure increases, and then, for very intense autocorrelations, start to decrease.
- This phenomenon, regarding the AR(1) process, vanishes for sample sizes greater than approximately 100, while, in contrast, the phenomenon remains in the FGN process case for very large samples.

## References


Feller, W. (1970), *An Introduction to Probability Theory and Its Applications, Vol. 2*, 2nd ed., Wiley.

Soong, T. T. (2004), *Fundamentals of Probability and Statistics for Engineers*, 1st ed., Wiley-Interscience.