Volatility Leverage ARCH Models with Non-Gaussian Shocks

PRESENTED BY

Krzysztof Podgórski

Lund University

WORK IS JOINTLY WITH

Farrukh Javed

Örebro University

November 11, 2019

Inspirational Anecdote

The French physicist Gabriel Lippman wrote the following in a letter to Henri Poincare.

Tout le monde y croit cependent, car les experimenteurs s'imaginent que c'est un theorem de mathematiques, et les mathematiciens que c'est un fait experimental.

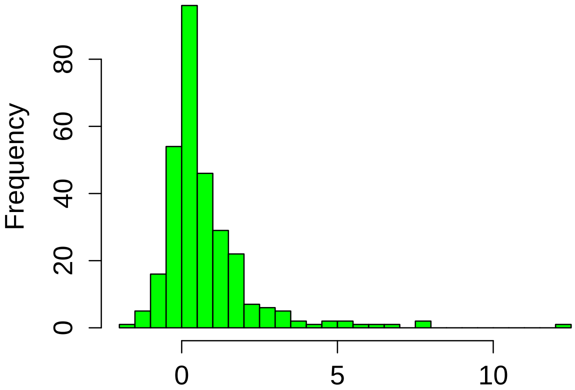
Everybody believes in the exponential law of errors [a]:

- *the experimenters, because they think it can be proved by mathematics;*
- *the mathematicians, because they believe it has been established by observation.*

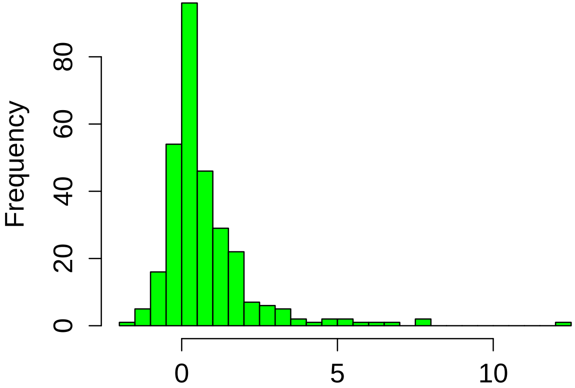
[a] exponential law refers to what nowadays we call Gaussian or normal distribution.

Outline

Good grief, data are heavy tailed, skewed!



Good grief, data are heavy tailed, skewed!



What should I do?!

Should one do anything?

- Got data $(y_1, \ldots, y_n)$.
- Want variance, take $\overline{(y - \bar{y})^2}$ and quote the law of large numbers.
- If one asks about the efficiency, call it *quasi maximum likelihood* estimator quoting the MLE for gaussian distribution.
- There is a good chance that some will be tricked into believing that 'quasi' refers to 'maximum' not to likelihood.
- Thus you can pretend that your estimator is sanctified by the efficiency of likelihood estimation methods.

Should one do anything?

- Got data $(y_1, \ldots, y_n)$.
- Want variance, take $\overline{(y-\bar{y})^2}$ and quote the law of large numbers.
- If one asks about the efficiency, call it *quasi maximum likelihood* estimator quoting the MLE for gaussian distribution.
- There is a good chance that some will be tricked into believing that 'quasi' refers to 'maximum' not to likelihood.
- Thus you can pretend that your estimator is sanctified by the efficiency of likelihood estimation methods.

It would not look so good if your method would be called *'maximum wrong likelihood'*, as it should be.

Why should we care?

Why should we care?

This little experiment tells why.

```
#Pareto
 alpha=2.1; u=runif(1000); y=u^(-1/alpha)

#Mean and variance
 mu=1/(1-1/alpha) ; s2=1/(1-2/alpha)-mu^2

#Maximum wrong likelihood estimators
 bary=mean(y); vary=var(y)

#'Not-wrong model' based estimation
 hatalpha=1/mean(log(y));
 hatmu=1/(1-1/hatalpha); hats2=1/(1-2/hatalpha)-hatmu^2
> mu               > bary            > hatmu
 1.90909             1.84928          1.91329
> sigma2           > vary            > hatsigma2
 17.3553            2.13148           18.4065
```

Why people do it?

- Model are complex and the Gaussian paradigm is build into some more complex often non-linear structure.
- The model driven data non necessarily are Gaussian.
- The Gaussianity of the underlying variables is not directly verifiable.
- The Gaussian likelihood method are efficient numerically.
- The robustness for deviation from the distributional assumptions is not sufficiently studied.
- Non-Gaussianity of residuals is often swept under the carpet of 'quasi likelihoodeness'.

Asymmetric Power ARCH model

- Ding, Granger and Engle (1993) in "A long memory property of stock market returns and a new model" addressed the leverage effect through the model

Asymmetric Power ARCH model

- Ding, Granger and Engle (1993) in "A long memory property of stock market returns and a new model" addressed the leverage effect through the model

$$y_t = m + a_1 y_{t-1} + \epsilon_t, \quad \epsilon_t = \rho_t e_t, \quad e_t \sim N(0,1)$$

$$\rho_t^\delta = \alpha_0 + \alpha_1 \, \rho_{t-1}^\delta \, [(1-\theta)^\delta e_{t-1}^{+\delta} + (1+\theta)^\delta e_{t-1}^{-\delta} + \beta],$$

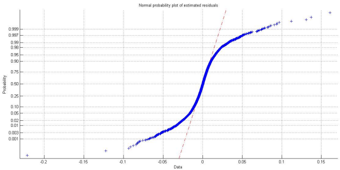
- The standardized errors e_t are Gaussian
- However, the non-standardized ones $\epsilon_t = \rho_t e_t$ are not due to presence random volatility ρ_t.
- If one just looks at regular residuals, then one should not be bothered by their non-Gaussianity.
- If one looks at standardized ones and they are not Gaussian, argument quasi-likelihood is often invoked.

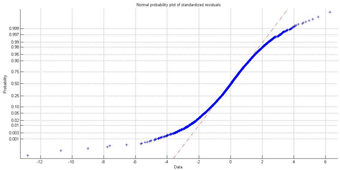
Analyzing the performance of model

- We fit the real data (S&P500) to the A-PARCH model and analyze the residuals (left) and standardized residuals (right).

Analyzing the performance of model

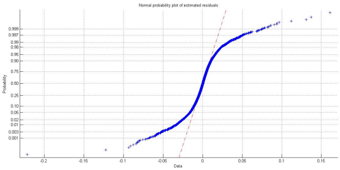
- We fit the real data (S&P500) to the A-PARCH model and analyze the residuals (left) and standardized residuals (right).

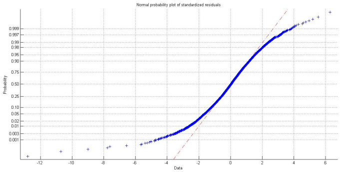




Analyzing the performance of model

- We fit the real data (S&P500) to the A-PARCH model and analyze the residuals (left) and standardized residuals (right).

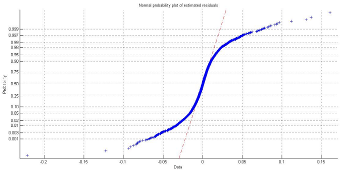


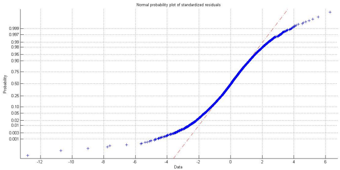


- Both the residuals ($\hat{\epsilon}_t$) and the standardized residuals ($\hat{e}_t$) are producing tails heavier than normal distribution.

Analyzing the performance of model

- We fit the real data (S&P500) to the A-PARCH model and analyze the residuals (left) and standardized residuals (right).

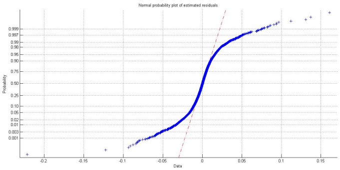


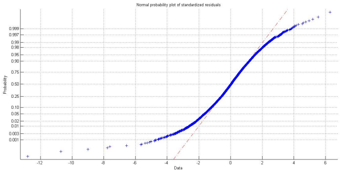


- Both the residuals ($\hat{\epsilon}_t$) and the standardized residuals ($\hat{e}_t$) are producing tails heavier than normal distribution.
- The model removed significant amount of mass from the tails.

Analyzing the performance of model

- We fit the real data (S&P500) to the A-PARCH model and analyze the residuals (left) and standardized residuals (right).





- Both the residuals ($\hat{\epsilon}_t$) and the standardized residuals ($\hat{e}_t$) are producing tails heavier than normal distribution.
- The model removed significant amount of mass from the tails.
- The standardized residuals still show excess kurtosis.

Challenges

- The standardized residuals still show excess kurtosis.

Challenges

- The standardized residuals still show excess kurtosis.
- The problem may not necessarily lie in the model as the Gaussian errors enter the model in non-linear fashion and deviation from normality maybe due to the slow rate of convergence.
- There is a need first to investigate if the Gaussian likelihood method will work for such non-linear models with the Gaussian error.

Challenges

- The standardized residuals still show excess kurtosis.
- The problem may not necessarily lie in the model as the Gaussian errors enter the model in non-linear fashion and deviation from normality maybe due to the slow rate of convergence.
- There is a need first to investigate if the Gaussian likelihood method will work for such non-linear models with the Gaussian error.
- Alternatively, the model with non-Gaussian errors may produce more efficient fit.

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Data	**Kurtosis**		
	Returns y_t	Residuals with volatility $\hat{\rho}_t \hat{e}_t$	Residuals $\hat{e}_t$
S&P500 data	26.12	24.89	8.18
Simulation	7.33	7.23	2.96

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Data	Kurtosis		
	Returns y_t	Residuals with volatility $\hat{\rho}_t \hat{e}_t$	Residuals $\hat{e}_t$
S&P500 data	26.12	24.89	8.18
Simulation	7.33	7.23	2.96

- Kurtoses for real data and estimated residuals are similar.

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Data	Kurtosis		
	Returns y_t	Residuals with volatility $\hat{\rho}_t \hat{e}_t$	Residuals $\hat{e}_t$
S&P500 data	26.12	24.89	8.18
Simulation	7.33	7.23	2.96

- Kurtoses for real data and estimated residuals are similar.
- Kurtosis for residuals is higher than 3 – value for normal distribution.

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Data	Kurtosis		
	Returns y_t	Residuals with volatility $\hat{\rho}_t \hat{e}_t$	Residuals $\hat{e}_t$
S&P500 data	26.12	24.89	8.18
Simulation	7.33	7.23	2.96

- Kurtoses for real data and estimated residuals are similar.
- Kurtosis for residuals is higher than 3 – value for normal distribution.
- Distribution of errors has tails heavier than normal.

Parametric boostrap study

- Simulate random data from the same model with Gaussian with the parameters estimated using the real data.

Data	Kurtosis		
	Returns y_t	Residuals with volatility $\hat{\rho}_t \hat{e}_t$	Residuals $\hat{e}_t$
S&P500 data	26.12	24.89	8.18
Simulation	7.33	7.23	2.96

- Kurtoses for real data and estimated residuals are similar.
- Kurtosis for residuals is higher than 3 – value for normal distribution.
- Distribution of errors has tails heavier than normal.
- The simulated data couldn't generate the magnitude of tails as seen in the data.

Non-Gaussian errors

- Clearly, the model needs 'non-Gaussian' enhancement to account for the observed deviations from the Gaussianity of standardized residuals.
- Even if one goes for **quasi MLE**, or (M-quasi-LE), the quality of the estimation needs non-Gaussian benchmarking.
- Preferably, the 'correct' MLE should be used but for this concrete and analitically trackable distributions for errors are needed.
- Inspired by:

Jensen, M. B. and Lunde, A. (2001).
The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model.
Econometrics Journal, **4**:167–342.

we investigate a general structure encompassing a number of possible non-Gaussian alternatives to the original AP-ARCH model.

Outline

Risk aversion

- In the GARCH in mean model for equities y_t with the volatility r_t, the risk averse investors are accounted for by $h(\cdot)$ in the following

$$y_t = f(y_{t-1}, \dots) + h(r_t) + r_t\epsilon_t,$$

where the form of f is not of interest for our discussion.
- For our approach, the specific form of h is not critically important.
- For the sake of simplicity we take $h(r) = r$. In the past research, various other risk aversion functions h were considered such as, for example, $h(r) = r^2$, or even $h(r) = \log r$
- The discussion would have to be adjusted for such a modification.
- Without losing generality we can assume the equity equation

$$y_t = c + dr_t + r_t\epsilon_t.$$

- The benefits of using the GARCH in mean is well discussed in the literature and the parameter d can be easily accounted for, thus in the following we set $d = 0$.

Mean-Variance Mixture of Gaussian

- Non-Gaussianity can start at the level of the equity equation
- To account for heavier tails and different asymmetries in the model we consider variance-mean Gaussian mixtures for distribution of ϵ_t's. Then the generic model can be written as follows

$$y_t = c + r_t \left(\mu \Gamma_t + \sqrt{\Gamma_t} e_t \right), \quad e_t \sim N(0, 1).$$

- Here the error $\epsilon_t = \mu \Gamma_t + \sqrt{\Gamma_t} e_t$ represents mixture of e_t by an independent mixture variable Γ_t.
- For this work the distribution of Γ_t, which determines both tails and asymmetries, is of interest.

Mean-Variance Mixture of Gaussian

- The generalized hyperbolic (GH) distributions have been successfully applied to various financial data and are fairly flexible while simple alternative to the Gaussian distributions
- They are characterized by their log-density being a hyperbola and considered as a reasonable alternative for modeling heavier than normal tailed data and known to simultanously capture the peaked and skewed behavior of returns data.
- They can be conveniently represented as a normal variance-mean mixture

$$X = \sigma\sqrt{\Gamma}Z + \mu\Gamma,$$

  where Z is a standard normal variable independent of Γ distributed according to generalized inverse Gaussian (GIG).
- The density of Γ :

$$f_\Gamma(x) = \frac{(a/b)^{\tau/2}}{2K_\tau(\sqrt{ab})} x^{\tau-1} e^{-\frac{ax+b/x}{2}},$$

  where $a > 0, b \geq 0$, if $\tau > 0$;
  $a > 0, b > 0$, if $\tau = 0$;
  and $a \geq 0, b > 0$, if $\tau < 0$,
  and K_τ is the modified Bessel function of the second kind.

Outline

General framework of the models

- While the **mean equity equation** involves non-Gaussian innovations, what is critical for the model and its capacity to produce the **leverage effect** is how these innovations enter the **volatility equation**.
- This, however, can be done in a number of different ways.
    - One can use only positive and negative parts of the Gaussian component e_t to desymmetrize volatility, the path taken for example in Jensen and Lunde.
    - This desymmetrization can be achieved by using the entire innovation $\epsilon_t = \mu\gamma_t + \sqrt{\gamma_t}e_t$.

  To be more general consider an autoregressive volatility model

$$r_t^\delta = \alpha_0 + r_{t-1}^\delta \lambda_{t-1}, \quad , t \in \mathbb{Z},$$

  in which λ_t is a function of some random components of the innovations ϵ_t.

Mathematical formulation

- Mathematically, if $\epsilon_t = u(\chi_t)$, where χ_t a random variable or vector while u is a given deterministic transformation, then $\lambda_t = v(\chi_t)$, for a certain other deterministic transformation v.
- Both u and v may depend on some parameters that maybe of interest.
- We illustrate this set-up through three important specifications of λ_t in the volatility equation

$$\chi_t = (\gamma_t, e_t);\;\; u(\gamma_t, e_t) = \mu\gamma_t + \sqrt{\gamma_t}e_t;\;\; v(\gamma_t, e_t) = \begin{cases} \alpha\left[(1-\theta)e_t^+ + (1+\theta)e_t^-\right] + \beta, \\ \alpha\left[(1-\theta)\epsilon_t^+ + (1+\theta)\epsilon_t^-\right] + \beta, \end{cases}$$

$$\chi_t = (\gamma_t^+, \gamma_t^-);\;\; u(\gamma_t^+, \gamma_t^-) = \kappa\gamma_t^+ - \gamma_t^-/\kappa;\;\; v(\gamma_t^+, \gamma_t^-) = \alpha\left[(1-\theta)\gamma_t^+ + (1+\theta)\gamma_t^-\right] + \beta.$$

- In the first case only the Gaussian component in a shock affects volatility directly,
- In the second we have the entire non-Gaussian shock involved. In this case, even when $\theta = 0$ one, in principle, may observe asymmetry in the response to the news attributed to the asymmetry in ϵ_t. In fact, it is interesting to investigate this effect in such a case, since avoiding positive and negative parts allows for analytically natural models.
- In the third case, the innovation itself is presented as a decomposition into positive and negative components γ_t^+, γ_t^- that are,
  for example, independently gamma distributed.

Benefits of the general formulations

- A lot of mathematical and statistical properties are shared with the Gaussian model that has been discussed in full detail in our previous work.

Benefits of the general formulations

- A lot of mathematical and statistical properties are shared with the Gaussian model that has been discussed in full detail in our previous work.
- The general treatment for the volatility equation allows to obtain results independently of the specific form of χ, u and v.
- Then pros and cons of different specifications of χ, u and v can be analyzed on the derived general properties.

Generalized Laplace alternative

Taking errors as gammma variance-mean mixture

$$e_t = \sqrt{\gamma}_t Z_t + \mu\gamma_t,$$

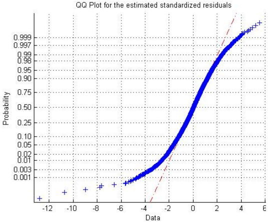
where γ_t are iid gamma distributed and Z_t iid standard normal, we obtain the model with generalized asymmetric Laplace distributions

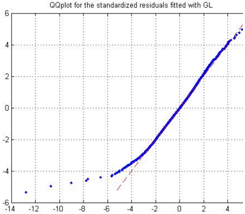
$$y_t = m + a\, y_{t-1} + \rho_t \left(\mu\gamma_t + \sigma\sqrt{\gamma_t} Z_t\right),$$

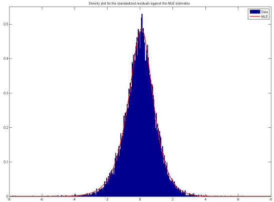
The relation for volatility is the same as in the original model i.e.

$$\rho_t^\delta = \alpha_0 + \alpha_1\ \rho_{t-1}^\delta\ [(1-\theta)^\delta e_{t-1}^{+\delta} + (1+\theta)^\delta e_{t-1}^{-\delta} + \beta],$$

Fit of residuals







Sample distribution of residuals.

- *Left*: qq-plot against the Gaussian distribution;
- *Middle*: qq-plot against the fitted GAL;
- *Right*: Histogram of the residuals against the fitted GAL density.

Alternative Model with GAL errors within the framework

An important property of $Y \sim \mathbb{GL}(m, \kappa, \sigma, \tau)$ is that, it can be characterized as a difference of two gamma distributed random variables

$$Y \overset{d}{=} m + \rho_t e_t$$

where the e_t follows GL distribution with the following representation

$$e_t = \kappa \gamma_t^+ - \frac{1}{\kappa} \gamma_t^-$$

The γ^+ and γ^- are the two independent and identically distributed gamma random variables with the same shape parameter τ

Volatility

$$\rho_t^\delta = \alpha_0 + \alpha \rho_{t-1}^\delta [(1-\theta)^\delta {\gamma_{t-1}^+}^\delta + (1+\theta)^\delta {\gamma_{t-1}^-}^\delta + \beta]$$

Outline

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.
- It is also essential to build a model with relatively small and meaningful parameters so that it can be interpretable across different economic environments.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.
- It is also essential to build a model with relatively small and meaningful parameters so that it can be interpretable across different economic environments.
- Our proposed models are capable of handling the distinct feature of financial data and generalize several different approaches both with respect structure and with respect distributional properties.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.
- It is also essential to build a model with relatively small and meaningful parameters so that it can be interpretable across different economic environments.
- Our proposed models are capable of handling the distinct feature of financial data and generalize several different approaches both with respect structure and with respect distributional properties.
- The difference of Γ's allows us to model asymmetry in a natural way avoid using absolute values of asymmetric random variables.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.
- It is also essential to build a model with relatively small and meaningful parameters so that it can be interpretable across different economic environments.
- Our proposed models are capable of handling the distinct feature of financial data and generalize several different approaches both with respect structure and with respect distributional properties.
- The difference of Γ's allows us to model asymmetry in a natural way avoid using absolute values of asymmetric random variables.
- The heavy tails/long memory are naturally modeled by the shape rather than the power that was shown not having the desired effect.

Conclusions

- Financial data exhibits certain features such as **heavy tail**,**asymmetry**, **leverage effect** and **volatility clustering**.
- In the studied CH volatility models the non-Gaussianity observed in the data could not be explained by non-linearity of the models.
- Non-Gaussian models are needed both for accuracy and for verification Gaussian likelihood based methods.
- It is also essential to build a model with relatively small and meaningful parameters so that it can be interpretable across different economic environments.
- Our proposed models are capable of handling the distinct feature of financial data and generalize several different approaches both with respect structure and with respect distributional properties.
- The difference of Γ's allows us to model asymmetry in a natural way avoid using absolute values of asymmetric random variables.
- The heavy tails/long memory are naturally modeled by the shape rather than the power that was shown not having the desired effect.
- More studies are needed to verify the models in applications.

Some final thoughts courtesy of George E. P. Box

Some final thoughts courtesy of George E. P. Box

- *Essentially, all models are wrong, but some are useful.*

Some final thoughts courtesy of George E. P. Box

- *Essentially, all models are wrong, but some are useful.*
- *Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.*

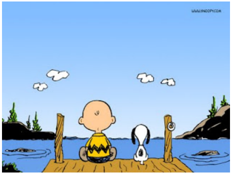
Some final thoughts courtesy of George E. P. Box

- *Essentially, all models are wrong, but some are useful.*
- *Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.*
- *A man in daily muddy contact with field experiments could not be expected to have much faith in any direct assumption of independently distributed normal errors.*

Some final thoughts courtesy of George E. P. Box

- *Essentially, all models are wrong, but some are useful.*
- *Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.*
- *A man in daily muddy contact with field experiments could not be expected to have much faith in any direct assumption of independently distributed normal errors.*

George E. P. Box – a British mathematician and Professor of Statistics at the University of Wisconsin



Thank you!