

The Fence around Least Squares Is Linearity, Not Normality

— Gauss's Two Theorems: 1809 Assumes an Error Law and 1821 Assumes None — [Paper 171]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

August 2026

— *From thinking to syncing.* —

ABSTRACT

Paper 157 wrote that the 2 in least squares is an assumed error law. This paper stands at the other fence — Gauss answered the same question twice, and the second answer (1821) assumes no distribution at all. And “best” is still fenced — by linearity, not by normality. **NO NEW MATHEMATICAL THEOREM IS CLAIMED.** First, Gauss is one person with two theorems. The 1809 argument assumes that the arithmetic mean is best and deduces that the error law is normal; the 1821 argument assumes only zero mean, equal variance and no correlation, and deduces minimum variance (Section 2). Second, Gauss–Markov was verified directly. Among 200000 random weight vectors summing to 1, not one fell below $1/n = 0.0909090909$ — the best was 0.0919952051. No distribution was used anywhere in that section (Section 3). Third, and this is the core. Even where the variance is finite and Gauss–Markov holds in full (a t distribution with 5 degrees of freedom), dropping linearity beats the mean — 0.8385 times its error for the trimmed mean and 0.8434 for the Huber estimator (Section 4). Fourth, yet “nonlinear therefore better” is false. In the same setting the median loses (1.0467 times) — leaving the linear class is necessary and not sufficient (Section 4). Fifth, unbiasedness and efficiency have distinct roots. With unequal variances ($\sigma_i^2 = i$) the plain mean and the inverse-variance weighting are both unbiased and differ in variance by 1.65 times (Section 5).

Scope of this paper: **NO NEW MATHEMATICAL THEOREM IS CLAIMED** — the Gauss–Markov theorem, the Cramér–Rao bound, M -estimation and generalised least squares are all standard. Nor is least squares criticised — what is shown is how far its “best” extends. On the overlap with Paper 157: 157 showed that the exponent 2 is an assumed error law, and this paper shows that the argument which assumes no error law is fenced too — two fences around one theorem, and nothing there is retracted. What is added is verifying Gauss–Markov directly by random search, showing with the median that “nonlinear therefore better” is false, and setting out in numbers that unbiasedness and efficiency can hold and fail independently.

1. Notation and position

An estimator is **linear** when it can be written as a weighted sum $\sum w_i x_i$ of the observations; **unbiased** when its expectation is the true value; **efficient** when its variance is small. Every number here was computed with home-made random numbers (normals by Box–Muller, the t distribution built from normals and a chi-square).

Position: Paper 157 gave the 2 of the square its address as an assumed error law. This paper stands at the fence next to it — the argument that assumes no error law has a fence of its own.

2. Gauss is one person with two theorems

Fact 2.1

work	assumption	conclusion
1809 Theoria Motus	that the mean is best	the error law is normal
1821 Theoria Combinationis	no distribution at all	minimum variance

Observation 2.2

The 1809 argument puts into its assumptions what one wants as its conclusion. Gauss himself replaced it in 1821 — the new argument demands only three things: zero mean, equal variances, no correlation.

Observation 2.3

“One person, two theorems” is a recurring type in this corpus (Yamabe, Hopf, Noether, Nash). Here too they are not a staircase of strength — the assumptions differ, and therefore so does how each breaks.

3. Verifying Gauss–Markov directly

Verification 3.1: With $n = 11$, generate 200000 random weight vectors summing to 1 and compare $\sum w_i^2$.

the uniform weights $1/n$	0.0909090909
best of 200000 random	0.0919952051
worst of 200000 random	0.2864219851

Not one fell below $1/n$. (QED)

Observation 3.2

No distribution was used in this section. All that was needed was equal variances and no correlation. The normal distribution appeared nowhere.

4. The core: the fence is linearity

Verification 4.1: Let the errors follow a t distribution with **5** degrees of freedom. Its variance is finite, so Gauss–Markov holds in full. Here $n = 11$ over **200000** trials.

estimator	linear?	bias	mean square error	ratio to the mean
mean (least squares)	yes	0.000577	0.15233893	1.000000
median	no	0.000948	0.15945434	1.046708
trimmed mean	no	0.000977	0.12773773	0.838510
Huber estimator	no	0.000836	0.12848074	0.843387

(QED)

Observation 4.2

The trimmed mean and the Huber estimator beat the mean. And Gauss–Markov is not violated — the mean is still best among linear estimators. The fence is linearity rather than normality, and it was in place before any distribution was mentioned.

Observation 4.3

But “nonlinear therefore better” is false — the median loses in the same setting (1.0467 times). Leaving the linear class is necessary and not sufficient. Being nonlinear buys the possibility of winning, not the win.

Verification 4.4: The same comparison with normal errors.

estimator	mean square error	ratio to the mean
mean	0.09078224	1.000000
median	0.13648426	1.503425
trimmed mean	0.10137188	1.116649

(QED)

Observation 4.5

With normal errors the mean wins. What normality does is promote the mean from best among linear estimators to best among all unbiased ones (Cramér–Rao). There are two “bests”, and only one of them needs an error law.

5. Unbiasedness and efficiency have distinct roots

Verification 5.1: Make the variances unequal, $\sigma_i^2 = i$ for $i = 1, \dots, 11$.

variance of the plain mean	0.5454545455
variance of the inverse-variance weighting	0.3311392768
ratio	1.647206
measured bias, plain mean	0.000952
measured bias, weighted	0.000498

(QED)

Observation 5.2

Both are unbiased, and their variances differ by 1.65 times. Unbiasedness and efficiency are separate properties, and what moves the second is the spread of the variances. The equal-variance clause of Gauss–Markov is not decoration: break it and the conclusion breaks — but unbiasedness does not.

6. And the 2 never reaches the answer

Verification 6.1: Minimising $\sum (y_i - a)^2$ gives $-2 \sum (y_i - a) = 0$, and the 2 divides out. Multiplying the square by any constant leaves the estimate untouched.

multiplier	estimate
1	0.700000000000
2	0.700000000000
7	0.700000000000
100	0.700000000000
the plain mean	0.700000000000

(QED)

Observation 6.2

The 2 of “squares” is not inside the estimator. What is inside is the exponent 2, which Paper 157 treated. They are two different twos.

7. The census

Observation 7.1

reading of “best”	theorem	what is assumed	the fence
least variance among linear unbiased	Gauss–Markov 1821	no distribution	linearity
least among all unbiased	Cramér–Rao	normality	the error law
maximum likelihood	one per error law	an error law	the mean if normal
robust	<i>M</i> -estimation, trimming	nonlinearity	wins on heavy tails

Observation 7.2

“Least squares is best” can be read four ways. Until it says which, the sentence has not yet become checkable (the same shape as what Paper 164 wrote about probability).

8. Conclusion

1. Gauss is one person with two theorems (Fact 2.1): **1809** assumes an error law, **1821** does not.
2. Gauss–Markov verified by random search (Ver. 3.1): **none of 200000** fell below $1/n$.
3. The fence is linearity (Ver. 4.1): **0.8385** trimmed, **0.8434** Huber.
4. “Nonlinear therefore better” is false (Obs. 4.3): **the median at 1.0467**.
5. Unbiasedness and efficiency are distinct (Ver. 5.1): **both unbiased, variance 1.65** times apart.
6. The **2** of the square never reaches the answer (Ver. 6.1).

Closing: There is no argument that assumes nothing. Gauss dropped the assumption of an error law in 1821, and in its place he laid down linearity — he simply did not write that he had. When told that nothing has been assumed, what to look for is not the absence of an assumption but the assumption that was not written down.

References

1. Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. Hamburg: Perthes und Besser.
2. Gauss, C. F. (1821). *Theoria combinationis observationum erroribus minimis obnoxiae*. Göttingen: Dieterich.
3. Markov, A. A. (1912). *Wahrscheinlichkeitsrechnung*. Leipzig: Teubner.
4. Aitken, A. C. (1936). *On least squares and linear combination of observations*. Proceedings of the Royal Society of Edinburgh, 55.
5. Huber, P. J. (1964). *Robust estimation of a location parameter*. Annals of Mathematical Statistics, 35(1).

6. Tukey, J. W. (1960). *A survey of sampling from contaminated distributions*. In: Contributions to Probability and Statistics. Stanford University Press.
7. Yamagishi, Y. (2026). *The Equals Sign Has Distinct Roots Too* [Paper 16], 3rd ed. Zenodo. 10.5281/zenodo.21600471
8. Yamagishi, Y. (2026). *Zero-Tension and the Inverse Problem* [Paper 89]. Zenodo. 10.5281/zenodo.21748664
9. Yamagishi, Y. (2026). *Convention and Fact* [Paper 95]. Zenodo. 10.5281/zenodo.21758621
10. Yamagishi, Y. (2026). *Extracting Signal from Noise* [Paper 106]. Zenodo. 10.5281/zenodo.21836984
11. Yamagishi, Y. (2026). *The Fisher Metric Is Unique Only on the Classical Side* [Paper 133]. Zenodo. 10.5281/zenodo.22080550
12. Yamagishi, Y. (2026). *A Census of the Twos* [Paper 157]. Zenodo. 10.5281/zenodo.22091931
13. Yamagishi, Y. (2026). *Probability Has Four Addresses* [Paper 164]. Zenodo. 10.5281/zenodo.22092815
14. Yamagishi, Y. (2026). *Cancelling the Differentials Gives -1 , Not $+1$* [Paper 170]. Zenodo. 10.5281/zenodo.22093731

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)