

A Modern Gauss–Markov Theorem?

Hansen (2022) and the subsequent debate

Miguel Salema

May 4, 2026

Econometrics Reading Group
University of Bern

Introduction

Hansen (2022)

Hansen claims that Gauss-Markov theorem (GMT) and Aitken theorem finite-sample variance lower bounds (LB) can be restated without explicitly restricting attention to linear estimators.

⇒ OLS/GLS are the best unbiased estimators (BUE). Drop the “L” from BLUE.

Pötscher and Preinerstorfer (2023, 2024): relevance?

Could a nonlinear unbiased estimator beat OLS/GLS in finite samples?

The answer depends almost entirely on the class of data-generating distributions. Hansen's results either:

- Merely re-state Aitken or GMT;
- Add much more restrictive assumptions.

Follow-up papers

Back to the basics

Classical setup

Fixed-design linear regression:

Assumption 1: $Y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + e_{n \times 1}, \quad k < n,$

Assumption 2: $\mathbb{E}[e] = 0,$

Assumption 3: $0 < \text{Var}(e) = \mathbb{E}[ee'] = \sigma^2 \Sigma < \infty,$

where $\text{rank}(X) = k$ and Σ is symmetric, $n \times n$, and positive definite.

Standard estimators:

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y = AY,$$

$$\hat{\beta}_{\text{GLS}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y.$$

Definition 1: Variance comparison

For vectors, “smaller variance” means Loewner order: $\text{Var}_F(\hat{\beta}) - \text{Var}_F(\tilde{\beta}) \succeq 0.$

Definition 2: distribution classes

Let F be the CDF of Y . Construct sets of distributions:

$\mathcal{F}_2 = \bigcup_{\Sigma \succ 0} \mathcal{F}_2(\Sigma)$. Set of joint distributions that satisfy A.1-3.

Fixed covariance shape

$\mathcal{F}_2(\Sigma) =$

$\{F : \mathbb{E}_F[Y] = X\beta, \text{Var}_F(Y) = \sigma^2 \Sigma\}.$

Independent sampling

$\mathcal{F}_2^* \subset \mathcal{F}_2$. Diagonal Σ .

(Heteroskedastic case)

Homoskedastic

$\mathcal{F}_2^0 \subset \mathcal{F}_2^*.$

$\Sigma = I_n.$

Finite-sample variance lower bounds

Theorem 1: Gauss–Markov

If $\hat{\beta}$ is (i) a linear estimator and (ii) is unbiased $\forall F \in \mathcal{F}_2(I_n)$, then

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'X)^{-1} = \text{Var}_F(\hat{\beta}_{\text{OLS}}), \quad \text{for every } F \in \mathcal{F}_2(I_n).$$

Theorem 2: Cramér-Rao

If $\hat{\beta}$ is (i) unbiased $\forall F \in \mathcal{F}_2^\phi = \mathcal{F}_2(I_n) \cap \{e \sim \mathcal{N}(0, I_n\sigma^2)\}$, then

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'X)^{-1} = \text{Var}_F(\hat{\beta}_{\text{OLS}}), \quad \text{for every } F \in \mathcal{F}_2^\phi.$$

Theorem 3: Aitken

If $\hat{\beta}$ is (i) a linear estimator and (ii) is unbiased $\forall F \in \mathcal{F}_2(\Sigma)$, then

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'\Sigma^{-1}X)^{-1} = \text{Var}_F(\hat{\beta}_{\text{GLS}}), \quad \text{for every } F \in \mathcal{F}_2(\Sigma).$$

Hansen (2022)

Hansen's main claim

Theorem 4: modern Aitken

If $\hat{\beta}$ is (i) unbiased $\forall F \in \mathcal{F}_2$, then,

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'\Sigma^{-1}X)^{-1}, \quad \text{for every } F \in \mathcal{F}_2. [\text{Or } \mathcal{F}_2(\Sigma).]$$

[Equivalent, more useful formulation in squared parenthesis. To see Modern GMT plug-in $\Sigma = I_n$.]

Implication (wrong, under D.3)

T.1: OLS is BLUE if errors are homoskedastic;

T.2: OLS is BUE if errors are homoskedastic and normal;

T.3: GLS is BLUE.

- Hansen (2022) **T.4:** GLS is BUE and OLS is BUE under homoskedasticity.

Definition 3: Best estimators (very informal; source: me)

BUE/BLUE only if unbiased requested for exact Σ up to σ^2 rescaling.

The devil really is in the detail

Hansen (2022) (correctly) re-writes alternative (worse...) theorems:

Theorem 1(b): alternative Gauss-Markov

If $\hat{\beta}$ is (i) a linear estimator and (ii) is unbiased $\forall F \in \mathcal{F}_2^*$, then

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'X)^{-1} = \text{Var}_F(\hat{\beta}_{\text{OLS}}), \quad \text{for every } F \in \mathcal{F}_2^0.$$

Theorem 3(b): alternative Aitken

If $\hat{\beta}$ is (i) a linear estimator and (ii) is unbiased $\forall F \in \mathcal{F}_2$, then

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'\Sigma^{-1}X)^{-1} = \text{Var}_F(\hat{\beta}_{\text{GLS}}), \quad \text{for every } F \in \mathcal{F}_2. [\text{Or } \mathcal{F}_2(\Sigma).]$$

Pötscher–Preinerstorfer (2024)

Getting hot

Abstract

We show that the theorems in Hansen (2021a) (the version accepted by Econometrica), except for one, are not new as they coincide with classical theorems like the good old Gauss-Markov or Aitken Theorem, respectively; the exceptional theorem is incorrect. Hansen (2021b) corrects this theorem. As a result, all theorems in the latter version coincide with the above mentioned classical theorems. Furthermore, we also show that the theorems in Hansen (2022) (the version published in Econometrica) either coincide with the classical theorems just mentioned, or contain extra assumptions that are alien to the Gauss-Markov or Aitken Theorem.

Hansen (2021a,b, 2022) contain several assertions from which he claims it would follow that the linearity condition can be dropped from the Gauss-Markov Theorem or from the Aitken Theorem. We show that this conclusion is unwarranted, as his assertions on which this conclusion rests turn out to be only (intransparent) reformulations of the classical Gauss-Markov or the classical Aitken Theorem. Theorem 5 in Hansen (2021a,b) (i.e., his 'modern Gauss-Markov Theorem') is nothing else than the good old(fashioned) Gauss-Markov Theorem (i.e., Theorem 3.1 above), just stated in a somewhat unusual and intransparent way. [While the word 'linear' does not appear in the formulation of Theorem 5 in Hansen (2021a,b), linearity of the estimators is introduced indirectly through a backdoor provided by the stricter unbiasedness condition.] While Theorem 6 in Hansen (2021a,b) thus turns out to be correct, it is certainly not new. Theorem 6 in Hansen (2021b) is a special case of his Theorem 5 for the location model, and thus is also not new; in contrast, Theorem 6 in Hansen (2021a) is a special case of Assertion 1. Example A.1 in

The present paper is mainly pedagogical in nature. In particular, the results will not come as a surprise to anyone well-versed in the theory of linear models and familiar with basic concepts of statistical decision theory, but – given the confusion introduced by Hansen (2021a,b, 2022) – the paper will benefit the econometrics community.

One important upshot of the present paper is that one should *not* follow Hansen's plea to drop the linearity condition in teaching the Gauss-Markov Theorem or the Aitken Theorem. Depending on which formulation of the Gauss-Markov Theorem one starts with (Theorem 3.1 or 3.2 given below), dropping linearity from the formulation of that theorem at best leads to a result equivalent to the usual Gauss-Markov Theorem, and at worst leads to an incorrect result. The same goes for the Aitken Theorem. Unfortunately, in heeding his own advice Hansen has included an incorrect formulation of the Gauss-Markov Theorem in the August 2021 version of his forthcoming text-book (Theorem 4.4. in Hansen (2021d)) available on his webpage for an

We next turn to the treatment of the Gauss-Markov Theorem in Hansen (2022): He starts with Theorem 1 (for linear estimators), which despite given the label Gauss-Markov, is *not* the Gauss-Markov Theorem, but a much weaker result relying on the unnecessarily restrictive assumption that the errors in the model are independent (and homoskedastic).²⁴ Such an independence assumption is superfluous in the classical Gauss-Markov Theorem. [Note that as long as only linear estimators are considered, requiring unbiasedness for all $F \in \mathbf{F}_2^*$, as is done in

The P&P reframing: two equivalent classical statements

- Delete “linear” ($\hat{\beta} = CY$) from T.3(b) keeps it correct. Hence T.4 is correct.
- Removing “linear” from T.3 makes it false.

Theorem 7: Global unbiasedness forces linearity

If $\hat{\beta}$ is an estimator (i.e., Borel-measurable function of Y), with

$$\mathbb{E}_F[\hat{\beta}] = \beta(F) \text{ for every } F \in \mathcal{F}_2 \implies \exists C \in \mathbb{R}^{k \times n} : \hat{\beta} = CY.$$

Implication (true, this time)

- Hansen’s T.4 applies to exactly the same estimator class as T.3 (Aitken).
- It’s a mere mathematical re-writing. Linearity does not appear in the statement; but it’s reintroduced through global unbiasedness.

Source: Theorem 3.4 of [Pötscher and Preinerstorfer \(2024\)](#). [Portnoy \(2022\)](#) independently established a closely related linearity result, but P&P emphasize that it gives only Lebesgue almost-everywhere linearity. Hansen’s model class allows discrete distributions. A Lebesgue-null exceptional set can carry positive probability under a discrete distribution. See Remark 3.6 in [Pötscher and Preinerstorfer \(2024\)](#).

Deleting “linear” gives two different statements

Statement A: False in general.

Unbiased only over the fixed model $\mathcal{F}_2(\Sigma)$; ask whether GLS is best among all such unbiased estimators.

Non-linear estimators
may improve variance
for some distributions

Statement B: True, but...

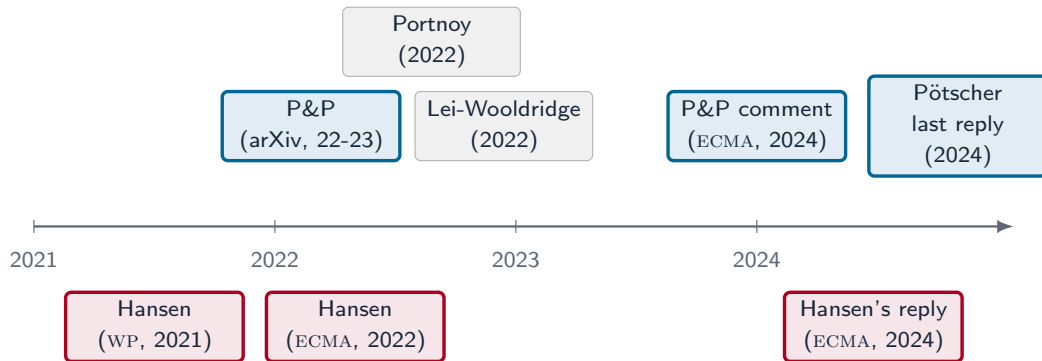
Unbiased over the union $\mathcal{F}_2 = \bigcup_{\Omega} \mathcal{F}_2(\Omega)$; evaluate bound under $\mathcal{F}_2(\Sigma)$.

Global unbiasedness
forces $\hat{\beta} = CY$

This is why I do not like the unqualified statement “GLS is BUE.” It is only BUE when we restrict the problem to exclude all competing nonlinear estimators!

Hansen (2024)

Timeline



Source: Timeline built from the references and chronology in [Pötscher and Preinerstorfer \(2023\)](#), [Hansen \(2024\)](#), and [Pötscher \(2024\)](#). This section actually presents results that were first published in [Hansen \(2022\)](#). But they came back again in the reply [Hansen \(2024\)](#), so I present them here.

Hansen's second (and better) claim

Theorem 4 allows arbitrary covariance/dependence, subject to finite second moments.

Theorem 5: modern Aitken for independent sampling.

If $\hat{\beta}$ is (i) unbiased $\forall F \in \mathcal{F}_2^*$, then,

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'\Sigma^{-1}X)^{-1}, \quad \text{for every } F \in \mathcal{F}_2^*. [\text{Or } \mathcal{F}_2^*(\Sigma).]$$

Theorem 6: modern Gauss-Markov for independent sampling.

If $\hat{\beta}$ is (i) unbiased $\forall F \in \mathcal{F}_2^*$, then,

$$\text{Var}_F(\hat{\beta}) \succeq \sigma^2(X'X)^{-1}, \quad \text{for every } F \in \mathcal{F}_2^0.$$

Implications (with a but...)

- Hansen (2022) **T.5**: GLS is BUE. (Not under D.3.)
- Hansen (2022) **T.6**: OLS is BUE under homoskedasticity. (Really not under D.3.)

Nonlinear unbiased estimators for Theorem 5?

Definition 4: Linear-plus-quadratic estimators

$$\hat{\beta}_{\alpha} = \hat{\beta}_{OLS} + \alpha \begin{bmatrix} Y' H_1 Y \\ \vdots \\ Y' H_k Y \end{bmatrix}, \quad \text{with } H_i = H_i', \operatorname{tr}(H_i \Sigma) = 0, X' H_i X = 0.$$

Lemma 1: Unbiased estimators for independent sampling

LPQ estimators are unbiased estimators for all F_2^* , independent sampling.

Source: Appendix A of [Pötscher and Preinerstorfer \(2023\)](#)

Unbiasedness class (columns) vs. evaluation class (rows)

	$\mathcal{F}_2(\Sigma)$	$\mathcal{F}_2$	$\mathcal{F}_2^*$	$\mathcal{F}_2^0$
$\mathcal{F}_2(\Sigma)$	NL unbiased can beat GLS. Linearity: Aitken $\Rightarrow$ GLS LB/BLUE.	GLS LB. Hansen T4 . P&P: unbiased $\Rightarrow$ linear. Aitken in disguise.	NL biased in general.	NL biased in general.
$\mathcal{F}_2^*(\Sigma)$	Over-strong, same as above.	Over-strong. Hansen T4 applies.	GLS LB. Hansen T5 . NL exists, LPS by Lemma 1 . So GLS is “BUE”.	NL biased in general.
$\mathcal{F}_2^0(I_n)$	Over-strong, same as above.	Over-strong. Hansen T4 applies.	OLS LB. Hansen T6 . NL unbiased exist. OLS is “BUE”.	NL unbiased can beat OLS. With linearity: GMT $\Rightarrow$ OLS LB/BLUE.

Notes: No independence is needed for GMT nor Aitken's. NL means nonlinear estimators; LB means finite-sample variance lower bound; LPQ denotes linear-plus-quadratic estimators.

Source: Me.

On Theorem 4

Hansen acknowledges that if all unbiased estimators are linear in that setting, this severely limits the relevance of Theorem 4.

On independent sampling

He shifts emphasis to Theorem 5: under independent sampling with arbitrary heteroskedastic variances, GLS is the best unbiased estimator; nonlinear unbiased estimators exist, but do not improve variance.

On teaching

He proposes replacing BLUE by BUE in the independent-sampling contexts, with careful explanations of what “unbiased” means.

Source: [Hansen \(2024\)](#).

Pötscher (2024), other critics, and conclusion

Getting even hotter

Abstract

In Pötscher and Preinerstorfer (2022) and in the abridged version Pötscher and Preinerstorfer (2024, published in *Econometrica*) we have tried to clear up the confusion introduced in Hansen (2022a) and in the earlier versions Hansen (201a,b). Unfortunately, Hansen's (2024) reply to Pötscher and Preinerstorfer (2024) further adds to the confusion. While we are already somewhat tired of the matter, for the sake of the econometrics community we feel compelled to provide clarification. We also add a comment on Portnoy (2023), a "correction" to Portnoy (2022), as well as on Lei and Wooldridge (2022).

We abstain from providing a list of all the errors and inconsistencies that can be found in Lei and Wooldridge (2022), and that make it difficult to gauge the actual contribution of that paper.

does establish linearity by exploiting unbiasedness under certain discrete distributions). It is welcomed that Portnoy now provides an alternative proof of the full linearity result, but this is an extension of Portnoy (2022) rather than a correction. Why this is advertised as a "correction" to Portnoy (2022) puzzles us. Also note that, for some reason, Portnoy (2023), and the supplementary material, do not acknowledge the full linearity result (Theorem 3.4) in Pötscher and Preinerstorfer (2022) at all.

A sketch of a proof of Theorem 11.1 is given in the book Hansen (2022b). This sketch is modelled after proofs in Hansen (2022a). Unfortunately, this sketch of a proof rests on Theorem 10.6 in Hansen (2022b), his rendition of the Cramér-Rao lower bound, a rendition that is lacking mathematical rigor and likely is not correct as given, because well-known regularity conditions (e.g., conditions for interchange of integration and differentiation) are omitted from Theorem 10.6 in Hansen (2022b). As a consequence of using Theorem 10.6 in Hansen (2022b), such regularity conditions are then not checked in the proof of Theorem 11.1. While the proof of Theorem 11.1 can perhaps be repaired, what is given in Hansen (2022b) certainly does not constitute a proof of Theorem 11.1. Fortunately, there is actually no need to come up with a proof of Theorem 11.1 as it is an old result already proved in Halmos (1946). This is well-known, and is also discussed in

The statement that Pötscher and Preinerstorfer (2024) does not examine the case of independent sampling is misleading to say the least: Hansen's (2022) results on independent sampling are discussed in the abstract, in the introduction, and in the conclusion of Pötscher and Preinerstorfer (2024), and the reader is referred to the more extensive discussion in Pötscher and Preinerstorfer (2022), a paper Hansen is well aware of. [This more extensive discussion was removed from Pötscher and Preinerstorfer (2024) on the request of the editor of *Econometrica*.]

Under independence, what changed?

Hansen's contribution

- In general α can reduce variance.
- However, $\alpha = 0$ is optimum for unbiased $\forall \mathcal{F}_2^*$.

The critics' message

- Classical GMT/Aitken don't require independence.
- T.5 has wider unbiasedness class than Aitken.
- T.5 is not that new: very similar to Koopmann (1982) and Gnot et al. (1992).
- T.6 is not useful.
- Language play more than mathematical contribution.
- Lei and Wooldridge (2022) get T.5 with restrictions on third moments rather than independence.
- Informal teaching statements risk being interpreted with fixed variances, which can make the statement false.

Source: Longer discussion in Pötscher and Preinerstorfer (2023) and Lei and Wooldridge (2022).

Thank you.

- Aitken, A. C. (1935). On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55, 42–48.
- Anderson, T. W. (1962). Least squares and best unbiased estimates. *Annals of Mathematical Statistics*, 33(1), 266–272.
- Fraser, D. (1954). Completeness of order statistics. *Canadian Journal of Mathematics*, 6, 42–45.
- Gnot, S., Knautz, H., Trenkler, G., and Zmyslony, R. (1992). Nonlinear unbiased estimation in linear models. *Statistics*, 23, 5–16.
- Halmos, P. R. (1946). The theory of unbiased estimation. *Annals of Mathematical Statistics*, 17(1), 34–43.
- Hansen, B. E. (2022). A modern Gauss–Markov theorem. *Econometrica*, 90(3), 1283–1294.
- Hansen, B. E. (2024). Reply to: Comment on “A Modern Gauss–Markov Theorem.” *Econometrica*, 92(3), 925–928.
- Koopmann, R. (1982). *Parameterschätzung bei a-priori-Information*. Vandenhoeck & Ruprecht.

- Lei, L. and Wooldridge, J. M. (2022). What estimators are unbiased for linear models? [arXiv:2212.14185](#).
- Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2), 99–135.
- Portnoy, S. (2022). Linearity of unbiased linear model estimators. *The American Statistician*, 76, 372–375.
- Portnoy, S. (2023). Correction to “Linearity of Unbiased Linear Model Estimators.” *The American Statistician*, 77, 237.
- Pötscher, B. M. (2024). Comments on B. Hansen’s Reply to A Comment on A Modern Gauss–Markov Theorem and some related discussion. [arXiv:2406.03971](#).
- Pötscher, B. M. and Preinerstorfer, D. (2023). A modern Gauss–Markov theorem? Really? [arXiv:2203.01425v5](#).
- Pötscher, B. M. and Preinerstorfer, D. (2024). A comment on “A Modern Gauss–Markov Theorem.” *Econometrica*, 92(3), 913–924.
- Rüschendorf, L. (1987). Unbiased estimation in nonparametric classes of distributions. *Statistics & Risk Modeling*, 5(1–2), 89–104.