



Munich Personal RePEc Archive

A Modern Gauss-Markov Theorem? Really?

Pötscher, Benedikt M. and Preinerstorfer, David

University of Vienna, University of St.Gallen

February 2022

Online at <https://mpra.ub.uni-muenchen.de/112185/>
MPRA Paper No. 112185, posted 08 Mar 2022 08:00 UTC

A Modern Gauss-Markov Theorem? Really?*

Benedikt M. Pötscher and David Preinerstorfer
Department of Statistics, University of Vienna
SEPS-SEW, University of St. Gallen

February 2022

Abstract

We show that the theorems in Hansen (2021b) (Econometrica, forthcoming) are not new as they coincide with classical theorems like the good old Gauss-Markov or Aitken Theorem, respectively.

1 Introduction

Hansen (2021b) contains 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, into which he has reintroduced linearity through the backdoor.

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 (2021b) – 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), 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 his forthcoming text-book (Theorem 4.4. in Hansen (2021a)).

*We would like to thank Abram Kagan for answering an inquiry and pointing us to the reference Kagan and Salaevskii (1969). Address correspondence to Benedikt Pötscher, Department of Statistics, University of Vienna, A-1090 Oskar-Morgenstern Platz 1. E-Mail: benedikt.poetscher@univie.ac.at.

2 The Framework

As in Hansen (2021b) we consider throughout the paper the linear regression model

$$Y = X\beta + e \quad (1)$$

where Y is of dimension $n \times 1$ and X is a (non-random) $n \times k$ design matrix with full column rank k satisfying $1 \leq k < n$.¹ It is assumed that

$$Ee = 0 \quad (2)$$

and

$$Eee' = \sigma^2 \Sigma, \quad (3)$$

where σ^2 , $0 < \sigma^2 < \infty$, is unknown and Σ is a known symmetric and positive definite $n \times n$ matrix ($Ee'e < \infty$). This model implies a distribution F for Y , which depends on X , β , and the distribution of e , in particular on σ^2 and Σ . Now define $\mathbf{F}_2(\Sigma)$ as the class of all such distributions F when β varies through \mathbb{R}^k and the distribution of e varies through all distributions compatible with (2) and (3) for the given Σ (and arbitrary σ^2 , $0 < \sigma^2 < \infty$). Of course, $\mathbf{F}_2(\Sigma)$ depends on X , but we suppress this in the notation. We furthermore introduce the set \mathbf{F}_2 as the larger class where we also vary Σ through the set of all symmetric and positive definite $n \times n$ matrices. In other words,

$$\mathbf{F}_2 = \bigcup_{\Sigma} \mathbf{F}_2(\Sigma),$$

where the union is taken over all symmetric and positive definite $n \times n$ matrices.² Again, \mathbf{F}_2 depends on X , but this dependence is not shown in the notation. The set \mathbf{F}_2^0 defined in Hansen (2021b) is nothing else than $\mathbf{F}_2(I_n)$, where I_n denotes the $n \times n$ identity matrix. In the following E_F (Var_F , respectively) will denote the expectation (variance-covariance matrix, respectively) taken under the distribution F . A word on notation: Given $F \in \mathbf{F}_2$, there is a unique β , denoted by $\beta(F)$, and a unique $\sigma^2 \Sigma$, denoted by $(\sigma^2 \Sigma)(F)$, compatible with the distribution F .

Remark 2.1. (*Ambiguity in the definition in Hansen (2021b)*) Hansen (2021b) also defines a set \mathbf{F}_2 , unfortunately somewhat ambiguously: Taking the first sentence mentioning his set \mathbf{F}_2 literally, his set would coincide with our $\mathbf{F}_2(\Sigma)$. The two sentences following that sentence, however, intimate that his set \mathbf{F}_2 was intended to coincide with our set \mathbf{F}_2 . This is confirmed by an inspection of his proofs; furthermore, if one would interpret his set \mathbf{F}_2 to mean our $\mathbf{F}_2(\Sigma)$, then the relation $\mathbf{F}_2^0 \subset \mathbf{F}_2$ given below (4) in Hansen (2021b) (which in our notation would become $\mathbf{F}_2(I_n) \subset \mathbf{F}_2(\Sigma)$) could not hold (except for Σ proportional to I_n). In the following we hence interpret Hansen's set \mathbf{F}_2 to coincide with our definition of \mathbf{F}_2 . In a remark further below

¹We make the assumption $k < n$ in order to use exactly the same framework as in Hansen (2021b).

²Note that $\mathbf{F}_2(\Sigma_1) \cap \mathbf{F}_2(\Sigma_2) = \emptyset$ iff Σ_1 and Σ_2 are not proportional. And $\mathbf{F}_2(\Sigma_1) = \mathbf{F}_2(\Sigma_2)$ iff Σ_1 and Σ_2 are proportional.

we discuss what happens if one would adopt the interpretation of Hansen's \mathbf{F}_2 as coinciding with our $\mathbf{F}_2(\Sigma)$.

3 The Gauss-Markov Case

To focus the discussion, we first treat the situation of a regression model with homoskedastic and uncorrelated errors, i.e., we assume that in (3) we have

$$\Sigma = I_n. \quad (4)$$

Let $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$ denote the ordinary least-squares estimator. The classical Gauss-Markov Theorem then reads as follows. Recall that a linear estimator is of the form AY , where A is a (nonrandom) $k \times n$ matrix. Also recall that $\mathbf{F}_2^0 = \mathbf{F}_2(I_n)$.

Theorem 3.1. *If $\hat{\beta}$ is a linear estimator that is unbiased under all $F \in \mathbf{F}_2^0$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2^0$), then*

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS})$$

for every $F \in \mathbf{F}_2^0$. [Here \succeq denotes Loewner order.]

The theorem can *equivalently* be stated in the following more unusual form, which is the form chosen by Hansen (see Theorem 1 in Hansen (2021b)).³

Theorem 3.2. *If $\hat{\beta}$ is a linear estimator that is unbiased under all $F \in \mathbf{F}_2$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), then*

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS}) \quad (5)$$

for every $F \in \mathbf{F}_2^0$.

In the latter theorem the unbiasedness is requested to hold over the *larger* class \mathbf{F}_2 of distributions rather than only over \mathbf{F}_2^0 . Of course, this is immaterial here and the two theorems are equivalent, because the estimators are required to be *linear* in both theorems and thus their expectation depends only on the first moment of Y and not on the second moments at all. While the difference in the unbiasedness conditions is immaterial in the preceding theorems, it is worth pointing out that the unbiasedness condition as given in Theorem 3.2 requires that an estimator is not only unbiased in the underlying model with uncorrelated and homoskedastic errors one is studying, but also requires unbiasedness under correlated and/or heteroskedastic errors (i.e.,

³As formulated in Hansen (2021b), his Theorem 1 has $\sigma^2(X'X)^{-1}$ instead of $Var_F(\hat{\beta}_{OLS})$ on the r.h.s. of the inequality. Taken literally this leaves σ^2 unspecified. To obtain a mathematically well-defined statement σ^2 needs to be interpreted as $\sigma^2(F)$, the variance of the data under F , the distribution under which the variance-covariance matrices of the estimators are computed.

under structures that are ‘outside’ of the model that is being considered). Why one would want to impose such a requirement when the underlying model has uncorrelated and homoskedastic errors is at least debatable. However, we stress once more that in the context of the preceding two theorems this does not matter.

We next discuss what happens if one eliminates the linearity condition in the two equivalent theorems. Dropping the linearity conditions leads to the following assertions, which will turn out to be *no longer* equivalent to each other:⁴

Assertion 1: If $\hat{\beta}$ is an estimator (i.e., a Borel-measurable function of Y) that is unbiased under all $F \in \mathbf{F}_2^0$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2^0$), then

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS})$$

for every $F \in \mathbf{F}_2^0$.

Assertion 2: If $\hat{\beta}$ is an estimator that is unbiased under all $F \in \mathbf{F}_2$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), then

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS})$$

for every $F \in \mathbf{F}_2^0$.

Not unexpectedly, Assertion 1 is *incorrect* in general.⁵ This is known. For the benefit of the reader we provide some counterexamples and attending discussion in the appendix. In particular, we see that in the classical Gauss-Markov Theorem *as it is usually formulated* (i.e., in Theorem 3.1) one can *not* eliminate the linearity condition in general!

Concerning Assertion 2, note that it coincides with Theorem 5 in Hansen (2021b) (his ‘modern Gauss-Markov Theorem’).⁶ Obvious questions now are (i) whether Assertion 2 (i.e., Theorem 5 in Hansen (2021b)) is correct, and (ii) if so, what is the reason for Assertion 2 to be correct while Assertion 1 is incorrect although in both assertions the linearity condition has been dropped. The answer to the latter question lies in the fact that Assertion 2 is requiring a *stricter* unbiasedness condition, namely unbiasedness over \mathbf{F}_2 rather than only unbiasedness over \mathbf{F}_2^0 . While the two unbiasedness conditions effectively coincide for *linear* estimators as discussed before, this is no longer the case once we leave the realm of linear estimators. Hence, the (potential) correctness of Assertion 2 (i.e., of Theorem 5 in Hansen (2021b)) must crucially rest on imposing the stricter unbiasedness condition, which not only requires unbiasedness under the model considered (regression with homoskedastic and uncorrelated errors), but oddly also under structures ‘outside’ of the maintained model (namely under heteroskedastic and/or correlated errors). Note that the class of competitors to $\hat{\beta}_{OLS}$ figuring in Assertion 1 is, in general, *larger* than the class of competitors appearing in Assertion 2.

⁴It is understood in both assertions that only estimators $\hat{\beta}$ are considered for which all appearing expressions $E_F \hat{\beta}$ and $Var_F(\hat{\beta})$ are well-defined.

⁵I.e., there exist design matrices X such that the assertion is false.

⁶The same caveat as expressed in Footnote 3 also applies to the formulation of Theorem 5 in Hansen (2021b).

Having understood what distinguishes Assertion 2 (i.e., Theorem 5 in Hansen (2021b)) from Assertion 1, the question remains whether the former is indeed correct, and if so, what its scope is, i.e., how much larger than the class of linear (unbiased) estimators the class of estimators covered by Assertion 2 (i.e., by Theorem 5 in Hansen (2021b)) is. We answer this now: As we shall show in the subsequent theorem, the only estimators $\hat{\beta}$ satisfying the unbiasedness condition of Assertion 2 (i.e., Theorem 5 in Hansen (2021b)) are *linear* estimators! *In other words, Hansen's Theorem 5 (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!* [Note that the word 'linear' does not appear in the *formulation* of Hansen's Theorem 5, but that linearity of the estimators is introduced indirectly through a backdoor provided by the stricter unbiasedness condition.] While Hansen's Theorem 5 thus turns out to be correct, it is certainly not new!⁷ Theorem 6 in Hansen (2021b) is a special case of his Theorem 5 for the location model, and thus is also not new.⁸ What has been said also serves as a reminder that one has to be careful with statements such as "best unbiased equals best linear unbiased". While this statement is incorrect in the context of Assertion 1 in general, it is trivially correct in the context of Assertion 2 (i.e., of Theorem 5 in Hansen (2021b)) as a consequence of the subsequent theorem.

An upshot of the preceding discussion is that – despite a plea to the contrary in Hansen (2021b) – one should **not** drop 'linearity' from the pedagogy of the Gauss-Markov Theorem. There is nothing to gain and a lot to lose: It will lead to an incorrect assertion, if one starts from the usual formulation of the classical Gauss-Markov Theorem (i.e., from Theorem 3.1); otherwise (i.e., if one starts from Theorem 3.2), it will lead to a correct, but rather intransparent, assertion that is in fact equivalent to the classical Gauss-Markov Theorem. Unfortunately, Hansen has fallen victim to his own advice as the Gauss-Markov Theorem (Theorem 4.4) given in his forthcoming text-book Hansen (2021a) is incorrect in general (as it coincides with Assertion 1).

Theorem 3.3. *If $\hat{\beta}$ is an estimator (i.e., a Borel-measurable function of Y) that is unbiased under all $F \in \mathbf{F}_2$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), then $\hat{\beta}$ is a linear estimator (i.e., $\hat{\beta} = AY$ for some $k \times n$ matrix A).^{9,10}*

We give a first "proof" based on Theorem 4.3 in Koopmann (1982) also reported as Theorem 2.1 in Gnot et al. (1992), but see the discussion below.

A first "proof": The unbiasedness assumption of the theorem obviously translates into

$$E_F \hat{\beta} = \beta(F) \text{ for every } F \in \mathbf{F}_2(\Sigma), \quad (6)$$

for *every* symmetric and positive definite Σ of dimension $n \times n$; specializing to the case $\Sigma = I_n$,

⁷We have not checked whether the *proofs* in Hansen (2021b) are correct or not.

⁸In contrast, Theorem 6 in the version of Hansen (2021b) available at the Econometrica website (i.e., Version September 2021) is a special case of Assertion 1. Example A.1 in the appendix shows that this theorem is false.

⁹By unbiasedness, such an A must then also satisfy $AX = I_n$.

¹⁰Curiously, the result by Koopmann (1982) in question is actually mentioned in Section 1 of Hansen (2021b).

we, in particular, obtain¹¹

$$E_F \hat{\beta} = \beta(F) \text{ for every } F \in \mathbf{F}_2(I_n). \quad (7)$$

Condition (7), together with Theorem 4.3. in Koopmann (1982) (see also Theorem 2.1 in Gnot et al. (1992)¹²), implies that $\hat{\beta}$ is of the form

$$\hat{\beta} = A^0 Y + (Y' H_1^0 Y, \dots, Y' H_k^0 Y)', \quad (8)$$

where A^0 satisfies $A^0 X = I_n$ and H_i^0 are matrices satisfying $\text{tr}(H_i^0) = 0$ and $X' H_i^0 X = 0$ for $i = 1, \dots, k$. It is easy to see that we may without loss of generality assume that the matrices H_i^0 are symmetric (otherwise replace H_i^0 by $(H_i^0 + H_i^{0'})/2$). Inserting (8) into (6) yields

$$E_F (A^0 Y + (Y' H_1^0 Y, \dots, Y' H_k^0 Y)') = \beta(F) \text{ for every } F \in \mathbf{F}_2(\Sigma),$$

and this has to hold for *every* symmetric and positive definite Σ . Standard calculations involving the trace operator and division by σ^2 now give

$$(\text{tr}(H_1^0 \Sigma), \dots, \text{tr}(H_k^0 \Sigma))' = 0 \text{ for every symmetric and positive definite } \Sigma. \quad (9)$$

For every $j = 1, \dots, n$, choose now a sequence of symmetric and positive definite matrices $\Sigma_m^{(j)}$ (each of dimension $n \times n$) that converges to $e_j(n) e_j(n)'$ as $m \rightarrow \infty$, where $e_j(n)$ denotes the j -th standard basis vector in \mathbb{R}^n (such sequences obviously exist). Plugging this sequence into (9), letting m go to infinity, and exploiting properties of the trace-operator, we obtain

$$(e_j(n)' H_1^0 e_j(n), \dots, e_j(n)' H_k^0 e_j(n))' = 0 \text{ for every } j = 1, \dots, n.$$

In other words, all the diagonal elements of H_i^0 are zero for every $i = 1, \dots, k$. Next, for every $j, l = 1, \dots, n$, $j \neq l$, choose a sequence of symmetric and positive definite matrices $\Sigma_m^{\{j,l\}}$ (each of dimension $n \times n$) that converges to $(e_j(n) + e_l(n))(e_j(n) + e_l(n))'$ as $m \rightarrow \infty$ (such sequences obviously exist). Then exactly the same argument as before delivers

$$((e_j(n) + e_l(n))' H_1^0 (e_j(n) + e_l(n)), \dots, (e_j(n) + e_l(n))' H_k^0 (e_j(n) + e_l(n)))' = 0 \text{ for every } j \neq l.$$

Recall that the matrices H_i^0 are symmetric. Together with the already established fact that the diagonal elements are all zero, we obtain that also all the off-diagonal elements in any of the matrices H_i^0 are zero; i.e., $H_i^0 = 0$ for every $i = 1, \dots, k$. This completes the proof.¹³ ■

¹¹Instead of I_n we could have chosen any other symmetric and positive definite $n \times n$ matrix Σ_0 instead.

¹²Note that X^- in that reference runs through all possible g -inverses of X .

¹³A slightly different version of the first "proof" can be obtained as follows. Theorem 4.3 in Koopmann (1982) shows for every given (fixed) Σ that any $\hat{\beta}$ satisfying (6) is of the form $AY + (Y' H_1 Y, \dots, Y' H_k Y)'$ where $AX = I_n$, the H_i 's satisfy $\text{tr}(H_i \Sigma) = 0$, and $X' H_i X = 0$ for $i = 1, \dots, k$. Again it is easy to see that we may assume the matrices H_i to be symmetric. Note that the matrices A and H_i flowing from Theorem 4.3 in Koopmann (1982)

Theorem 4.3. in Koopmann (1982) is proved by reducing it to Theorem 3.1 (via Theorems 3.2, 4.1, and 4.2) in the same reference. Unfortunately, a full proof of Theorem 3.1 is not provided in Koopmann (1982), only a very rough outline is given. Thus the status of Theorem 4.3 in Koopmann (1982) is not entirely clear. For this reason we next give a direct proof which does not rely on any result in Koopmann (1982).¹⁴

A direct proof: For every $m \in \mathbb{N}$, every $V = (v_1, \dots, v_m) \in \mathbb{R}^{n \times m}$ and $\alpha \in (0, 1)^m$ such that $\sum_{i=1}^m \alpha_i = 1$, define a probability measure via

$$\mu_{V, \alpha} := \sum_{i=1}^m \alpha_i \delta_{v_i},$$

where δ_z denotes unit point mass at $z \in \mathbb{R}^n$. The expectation of $\mu_{V, \alpha}$ equals $V\alpha$, and its covariance matrix equals $V \text{diag}(\alpha) V' - (V\alpha)(V\alpha)'$. Denote the expectation operator w.r.t. $\mu_{V, \alpha}$ by $E_{V, \alpha}$. Note that in case $V\alpha = 0$ and $\text{rank}(V) = n$ the measure $\mu_{V, \alpha}$ has expectation zero and a positive definite variance-covariance matrix; thus, $\mu_{V, \alpha}$ corresponds to an $F \in \mathbf{F}_2$ which has $\beta(F) = 0$. From the unbiasedness assumption imposed on $\hat{\beta}$ we obtain that

$$V\alpha = 0 \text{ and } \text{rank}(V) = n \text{ implies } 0 = E_{V, \alpha}(\hat{\beta}) = \sum_{i=1}^m \alpha_i \hat{\beta}(v_i). \quad (10)$$

Step 1: Fix $z \in \mathbb{R}^n$ and define $\alpha^{(1)} = 2^{-1}(n^{-1}, \dots, n^{-1})' \in \mathbb{R}^{2n}$, $\alpha^{(2)} = 2^{-1}((n+1)^{-1}, \dots, (n+1)^{-1})' \in \mathbb{R}^{2(n+1)}$, $V_1 = (I_n, -I_n)$ and $V_2 = (I_n, -I_n, z, -z)$. Clearly $V_1 \alpha^{(1)} = V_2 \alpha^{(2)} = 0$ and $\text{rank}(V_1) = \text{rank}(V_2) = n$. Furthermore,

$$\mu_{V_2, \alpha^{(2)}} = \frac{n}{n+1} \mu_{V_1, \alpha^{(1)}} + \frac{1}{2(n+1)} (\delta_z + \delta_{-z}). \quad (11)$$

Applying (10) with $E_{V_2, \alpha^{(2)}}$ and $E_{V_1, \alpha^{(1)}}$ now yields $0 = \hat{\beta}(z) + \hat{\beta}(-z)$, i.e., we have shown that

$$\hat{\beta}(-z) = -\hat{\beta}(z) \text{ for every } z \in \mathbb{R}^n, \quad (12)$$

in particular $\hat{\beta}(0) = 0$ follows.

Step 2: Let y and z be elements of \mathbb{R}^n . Define the matrix

$$A(y, z) = ((y_1 + z_1)e_1(n), \dots, (y_n + z_n)e_n(n)),$$

in principle could depend on Σ . The following argument shows that this is, however, not the case: If $\hat{\beta}$ had two distinct linear-quadratic representations, then the difference of these two representations would be a vector of multivariate polynomials (at least one of which is nontrivial) that would have to vanish everywhere, which is impossible since the zero-set of a nontrivial multivariate polynomial is a Lebesgue null-set. Given now the independence (from Σ) of the matrices H_i , one can then exploit the before mentioned relations $\text{tr}(H_i \Sigma) = 0$ in the same way as is done following (9) in the main text.

¹⁴Alternatively, one could try to provide a complete proof of the result in Koopmann (1982). We have not pursued this, but have chosen the route via a direct proof of Theorem 3.3.

where $e_i(n)$ denotes the i -th standard basis vector in \mathbb{R}^n , and set

$$V = (A(y, z), -y, -z, I_n, -I_n) \quad \text{and} \quad \alpha = (3n+2)^{-1}(1, \dots, 1)' \in \mathbb{R}^{3n+2}.$$

Then, we obtain $V\alpha = 0$ and $\text{rank}(V) = n$. Using (10) and (12) it follows that

$$0 = \sum_{i=1}^n \hat{\beta}((y_i + z_i)e_i(n)) + \hat{\beta}(-y) + \hat{\beta}(-z),$$

which by (12) is equivalent to

$$\hat{\beta}(y) + \hat{\beta}(z) = \sum_{i=1}^n \hat{\beta}((y_i + z_i)e_i(n)). \quad (13)$$

Using (13) with y replaced by $y + z$ and z replaced by 0 yields

$$\hat{\beta}(y + z) + \hat{\beta}(0) = \sum_{i=1}^n \hat{\beta}((y_i + z_i)e_i(n)).$$

Since $\hat{\beta}(0) = 0$ as shown before, we obtain

$$\hat{\beta}(y) + \hat{\beta}(z) = \hat{\beta}(y + z) \quad \text{for every } y \text{ and } z \text{ in } \mathbb{R}^n. \quad (14)$$

That is, we have shown that $\hat{\beta}$ is additive, i.e., is a group homomorphism between the additive groups \mathbb{R}^n and \mathbb{R}^k . By assumption it is also Borel-measurable. It then follows by a result due to Banach and Pettis (e.g., Theorem 2.2 in Rosendal (2009)) that $\hat{\beta}$ is also continuous. Homogeneity of $\hat{\beta}$ now follows from a standard argument, dating back to Cauchy, so that $\hat{\beta}$ is in fact linear. We give the details for the convenience of the reader: Relation (14) (which contains (12) as a special case) implies $\hat{\beta}(lz) = l\hat{\beta}(z)$ for every integer l . Replacing z by z/l ($l \neq 0$) in the latter relation gives $\hat{\beta}(z)/l = \hat{\beta}(z/l)$ for integer $l \neq 0$. It immediately follows that $\hat{\beta}(pz/q) = (p/q)\hat{\beta}(z)$ for every pair of integers p and q ($q \neq 0$). Let $c \in \mathbb{R}$ be arbitrary. Choose a sequence of rational numbers c_s that converges to c . Then by continuity of $\hat{\beta}$

$$\hat{\beta}(cz) = \lim_{s \rightarrow \infty} \hat{\beta}(c_s z) = \lim_{s \rightarrow \infty} (c_s \hat{\beta}(z)) = \left(\lim_{s \rightarrow \infty} c_s \right) \hat{\beta}(z) = c\hat{\beta}(z).$$

This concludes the proof. ■

Remark 3.4. (*Ambiguity in the definition in Hansen (2021b) continued*) If Hansen's \mathbf{F}_2 would be interpreted as coinciding with our $\mathbf{F}_2(\Sigma)$ (here with $\Sigma = I_n$ because of (4)) then the formulations of Theorems 3.1 and 3.2 as well as the formulations of Assertions 1 and 2 would coincide. In particular, with such an interpretation of Hansen's \mathbf{F}_2 his Theorem 5 would be false.

4 The Aitken Case

In this section we drop the assumption (4), i.e., Σ in (3) need not be the identity matrix. We make a preparatory remark: Similarly to observations made in Section 3 (see Footnote 3), the rendition of Aitken's Theorem (for linear estimators) as given in Theorem 3 in Hansen (2021b) needs some interpretation to convert it into a mathematically well-defined statement: The product $\sigma^2\Sigma$, on which the r.h.s. of the inequality in that theorem depends, is unspecified (note that σ^2 and Σ enter the expression only via the product), and needs to be interpreted as $(\sigma^2\Sigma)(F)$, the variance-covariance matrix of the data under the relevant F w.r.t. which the variance-covariances in this inequality are taken. The same comment applies to Theorem 4 in Hansen (2021b).

Aitken's Theorem as usually given in the literature reads as follows. Let $\hat{\beta}_{GLS} = \hat{\beta}_{GLS}(\Sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$ denote the generalized least-squares estimator using the known matrix Σ . Linear estimators are of the form $\hat{\beta} = AY$ where A is a (nonrandom) $n \times k$ matrix.

Theorem 4.1. *Let Σ be an arbitrary known symmetric and positive definite $n \times n$ matrix. If $\hat{\beta}$ is a linear estimator that is unbiased under all $F \in \mathbf{F}_2(\Sigma)$ (meaning that $E_F\hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2(\Sigma)$), then*

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{GLS})$$

for every $F \in \mathbf{F}_2(\Sigma)$.

Similar as in Section 3, due to linearity of the estimators, an *equivalent* version of the theorem is obtained if the unbiasedness requirement is extended to all of \mathbf{F}_2 . This is precisely what happens in Theorem 3 in Hansen (2021b), his rendition of the Aitken Theorem (for linear estimators). Note that the subsequent theorem is obviously equivalent to Theorem 3 in Hansen (2021b) and perhaps is more transparent. [To see the equivalence, note that the all-quantor over Σ in Theorem 4.2 can be "absorbed" by replacing $\mathbf{F}_2(\Sigma)$ with \mathbf{F}_2 provided the quantity $\sigma^2\Sigma$ appearing in the expression $Var_F(\hat{\beta}_{GLS}) = \sigma^2(X'\Sigma^{-1}X)^{-1} = (X'(\sigma^2\Sigma)^{-1}X)^{-1}$ in (15) below is understood as $(\sigma^2\Sigma)(F)$, as is necessary anyways for Theorem 3 in Hansen (2021b) to formally make sense as noted earlier.]

Theorem 4.2. *Let Σ be an arbitrary known symmetric and positive definite $n \times n$ matrix. If $\hat{\beta}$ is a linear estimator that is unbiased under all $F \in \mathbf{F}_2$ (meaning that $E_F\hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), then*

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{GLS}) \tag{15}$$

for every $F \in \mathbf{F}_2(\Sigma)$.

Dropping linearity in both theorems now leads to two assertions.

Assertion 3: Let Σ be an arbitrary known symmetric and positive definite $n \times n$ matrix. If $\hat{\beta}$ is an estimator that is unbiased under all $F \in \mathbf{F}_2(\Sigma)$ (meaning that $E_F\hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2(\Sigma)$), then

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{GLS})$$

for every $F \in \mathbf{F}_2(\Sigma)$.

Assertion 4: Let Σ be an arbitrary known symmetric and positive definite $n \times n$ matrix. If $\hat{\beta}$ is an estimator that is unbiased under all $F \in \mathbf{F}_2$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), then

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{GLS})$$

for every $F \in \mathbf{F}_2(\Sigma)$.

Assertion 3 is again incorrect in general for reasons similar to the ones given for Assertion 1 in the previous section. Assertion 4 is equivalent to Theorem 4 in Hansen (2021b) (his ‘modern Aitken Theorem’); this is seen in the same way as the equivalence of Theorem 4.2 above with Theorem 3 in Hansen (2021b). Assertion 4 is indeed correct, but again not new, as the class of estimators figuring in Assertion 4 coincides with the class of linear estimators as a consequence of Theorem 3.3 above. Furthermore, a comment like Remark 3.4 also applies here. We conclude this section by noting that the rendition of Aitken’s Theorem in the text-book Hansen (2021a) (Theorem 4.5) is ambiguously formulated, making it difficult to decide whether it coincides with the (incorrect) Assertion 3 or with Assertion 4, which is (trivially) correct.

5 Independent Identically Distributed Errors

We round-off the discussion by briefly considering in this section what happens if we add the condition

$$e_1, \dots, e_n \text{ are i.i.d.} \tag{16}$$

to the model where e_i denotes the i -th component of e . Let \mathbf{F}_2^{iid} be the subset of \mathbf{F}_2^0 corresponding to distributions F that result from (1), (2), (3), and (16). In particular, we ask what is the status of the following assertion which is analogous to Assertion 1.

Assertion 5: If $\hat{\beta}$ is an estimator that is unbiased under all $F \in \mathbf{F}_2^{iid}$ (meaning that $E_F \hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2^{iid}$), then

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS})$$

for every $F \in \mathbf{F}_2^{iid}$.

Note that Assertion 5 differs from Assertion 1 in two respects: (i) the set of competitors to $\hat{\beta}_{OLS}$, i.e., the set of unbiased estimators in Assertion 5 is potentially larger than the corresponding set in Assertion 1, and (ii) the set of distributions F for which the variance inequality has to hold has gotten smaller compared to Assertion 2. Hence, the truth-status of Assertion 1 does not inform us about the corresponding status of Assertion 5.

Fortunately, Example A.2 in the appendix comes to the rescue and shows that Assertion 5 is incorrect in general (meaning that a design matrix can be found such that it is false). This is so since the nonlinear estimator constructed in that example is a fortiori unbiased under \mathbf{F}_2^{iid} ,

and since the offending F found in that example in fact belongs to \mathbf{F}_2^{iid} . However, in the special case of the location model Assertion 5 is actually true. This follows directly from Theorem 5 in Halmos (1946). [Recall that, in contrast, Assertion 1 is false in the case of a location model; cf. Example A.1 in the appendix.]

If one restricts the distributions to the set $\mathbf{F}_2^{iid,ac} \subseteq \mathbf{F}_2^{iid}$ of distributions such that errors are not only i.i.d. but also are absolutely continuous, then in the location case the corresponding analogon to Assertion 5 is also true, see Example 4.2 in Section 2.4 of Lehmann and Casella (1988).

A nice result is due to Kagan and Salaevskii (1969): Suppose we restrict to i.i.d. errors in our regression model, but where now the distribution of the errors, G say, is known. Suppose also that $n \geq 2k + 1$ and that the design matrix has no rows of zeroes. Then, if $\hat{\beta}_{OLS}$ is best unbiased in this model, the distribution G must be Gaussian. [Kagan and Salaevskii (1969) actually prove a more general result.] A related result for the location model with independent (not necessarily identically distributed) errors is given in Theorem 7.4.1 of Kagan et al. (1973).

There is probably more in the mathematical statistics literature we are not aware of, but this is what a quick search has turned up.

A Appendix: Counterexamples

Here we provide various counterexamples to Assertion 1. They all rest on the following lemma which certainly is not original as similar computations can be found in the literature, see, e.g., Gnot et al. (1992) and references therein. Counterexamples can also be easily derived from results in the before mentioned papers. In this appendix we always maintain the model from Section 2 and assume that (4) holds.

Lemma A.1. *Consider the model as in Section 2, additionally satisfying (4).*

(a) *Define estimators via*

$$\hat{\beta}_\alpha = \hat{\beta}_{OLS} + \alpha(Y'H_1Y, \dots, Y'H_kY)' \quad (17)$$

where the H_i 's are symmetric $n \times n$ matrices and α is a real number. Suppose $\text{tr}(H_i) = 0$ and $X'H_iX = 0$ for $i = 1, \dots, k$. Then $E_F(\hat{\beta}_\alpha) = \beta(F)$ for all $F \in \mathbf{F}_2^0$.

(b) *Suppose the H_i 's are as in Part (a). If $\text{Cov}_F(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \dots, Y'H_kY)') \neq 0$ for some $c \in \mathbb{R}^k$ and for some $F \in \mathbf{F}_2^0$, then there exist an $\alpha \in \mathbb{R}^n$ such that*

$$\text{Var}_F(c'\hat{\beta}_\alpha) < \text{Var}_F(c'\hat{\beta}_{OLS}); \quad (18)$$

in particular, $\hat{\beta}_{OLS}$ then does not have smallest variance-covariance matrix (w.r.t. Loewner order) in the class of all estimators that are unbiased under all $F \in \mathbf{F}_2^0$.

(c) *Suppose the H_i 's are as in Part (a). For every $c \in \mathbb{R}^k$ and for every $F \in \mathbf{F}_2^0$ under which*

$\beta(F) = 0$ we have

$$Cov_F \left(c' \hat{\beta}_{OLS}, c'(Y' H_1 Y, \dots, Y' H_k Y)' \right) = \sum_{j=1}^n \sum_{l=1}^n \sum_{m=1}^n d_j \left(\sum_{i=1}^k c_i h_{lm}(i) \right) E(e_j e_l e_m), \quad (19)$$

where $d = (d_1, \dots, d_n)' = X(X'X)^{-1}c$ and $h_{lm}(i)$ denotes the (l, m) -th element of H_i .

(d) Suppose the H_i 's are as in Part (a). For every $c \in \mathbb{R}^k$ and for every $F \in \mathbf{F}_2^0$ under which (i) $\beta(F) = 0$ and under which (ii) the coordinates of Y are independent (equivalently, the errors e_i are independent)

$$Cov_F \left(c' \hat{\beta}_{OLS}, c'(Y' H_1 Y, \dots, Y' H_k Y)' \right) = \sum_{j=1}^n d_j \left(\sum_{i=1}^k c_i h_{jj}(i) \right) E(e_j^3). \quad (20)$$

Proof: The proof of Parts (a), (c), and (d) is by straightforward computation. Since

$$\begin{aligned} Var_F(c' \hat{\beta}_\alpha) &= Var_F(c' \hat{\beta}_{OLS}) + 2\alpha Cov_F \left(c' \hat{\beta}_{OLS}, c'(Y' H_1 Y, \dots, Y' H_k Y)' \right) \\ &\quad + \alpha^2 Var_F(c'(Y' H_1 Y, \dots, Y' H_k Y)'), \end{aligned}$$

the claim in (b) follows immediately as the first derivative of $Var_F(c' \hat{\beta}_\alpha)$ w.r.t. α and evaluated at $\alpha = 0$ equals $2Cov_F \left(c' \hat{\beta}_{OLS}, c'(Y' H_1 Y, \dots, Y' H_k Y)' \right)$. Hence, whenever this covariance is non-zero, we may choose $\alpha \neq 0$ small enough such that (18) holds. ■

We now provide a few counterexamples that make use of the preceding lemma.

Example A.1. Consider the location model, i.e., the case where $k = 1$ and $X = (1, \dots, 1)'$. Choose H_1 as the $n \times n$ matrix which has $h_{11}(1) = -h_{22}(1) = 1$ and $h_{ij}(1) = 0$ else. Then the conditions on H_1 in Part (a) of Lemma A.1 are satisfied, and hence $\hat{\beta}_\alpha$ is unbiased under all $F \in \mathbf{F}_2^0$. Setting $c = 1$, we find for the covariance in (20)

$$n^{-1}(E_F(e_1^3) - E_F(e_2^3)) \neq 0$$

for every $F \in \mathbf{F}_2^0$ under which $\beta(F) = 0$, the errors e_i are independent, and $E_F(e_1^3) \neq E_F(e_2^3)$ hold. Such distributions F obviously exist.¹⁵ As a consequence, $\hat{\beta}_{OLS}$ is not best unbiased in the class of all estimators $\hat{\beta}$ that are unbiased under all $F \in \mathbf{F}_2^0$. In particular, Assertion 1 is false for this design matrix.

For the argument underlying the preceding example it is key that the errors are *not* i.i.d. under the relevant F . In fact, in the location model (i.e., $X = (1, \dots, 1)'$) we have $Var_F(\hat{\beta}_{OLS}) \preceq Var_F(\hat{\beta}_\alpha)$ for every real α , for every choice of H_1 as in Part (a) of Lemma A.1, and for every $F \in \mathbf{F}_2^0$ under which the errors e_i are i.i.d., since then $Cov_F \left(\hat{\beta}_{OLS}, Y' H_1 Y \right) = 0$ as is easily seen. For other design matrices X the argument, however, works even for i.i.d. errors as we show in the subsequent example. Cf. Section 4.1 of Gnot et al. (1992) for related results and more.

¹⁵E.g., choose e_2, \dots, e_n i.i.d. $N(0, \sigma^2)$ and e_1 independent from e_2, \dots, e_n with mean zero, variance σ^2 , and third moment not equal to zero.

Example A.2. Consider the balanced one-way layout for $k = 2$ and $n = 4$. That is, X has first column equal to $(1, 1, 0, 0)'$ and second column equal to $(0, 0, 1, 1)'$. Set $c = (1, 0)'$. Then $d = (1/2, 1/2, 0, 0)'$. Choose, e.g., $H_1 = H_2$ as the 4×4 matrix made up of 2×2 blocks, where the off-diagonal blocks are zero, the first and second diagonal block, respectively, are given by

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then the conditions on H_i in Part (a) of Lemma A.1 are satisfied, and hence $\hat{\beta}_\alpha$ is unbiased under all $F \in \mathbf{F}_2^0$. For the covariance in (20) we find

$$(E_F(e_1^3) + E_F(e_2^3))/2$$

under any $F \in \mathbf{F}_2^0$ under which $\beta(F) = 0$ and the errors e_i are independent. If F is chosen such that the errors are furthermore i.i.d. and asymmetrically distributed, the expression in the preceding display reduces to $E_F(e_1^3) \neq 0$. Such distributions F obviously exist. As a consequence, $\hat{\beta}_{OLS}$ is not best unbiased in the class of all estimators $\hat{\beta}$ that are unbiased under all $F \in \mathbf{F}_2^0$. In particular, Assertion 1 is false for this design matrix.

Many more counterexamples can be generated with the help of Lemma A.1 as outlined in the subsequent remark.

Remark A.2. (i) Suppose X admits a choice of H_i satisfying the conditions in Part (a) of Lemma A.1 and a $c \in \mathbb{R}^k$ such that $\sum_{j=1}^n d_j \sum_{i=1}^k c_i h_{jj}(i) \neq 0$. Then the covariance in (20) is not zero if F in Part (d) of the lemma is chosen to correspond to asymmetrically distributed i.i.d. errors. Part (b) of the lemma can then be applied. In case $H_i = H$ for all $i = 1, \dots, k$, these conditions further reduce to $\sum_{j=1}^n d_j h_{jj} \neq 0$ and $\sum_{i=1}^k c_i \neq 0$.

(ii) Suppose X admits a choice of H_i satisfying the conditions in Part (a) of Lemma A.1 and a $c \in \mathbb{R}^k$ such that for an index j_0 it holds that $d_{j_0} \sum_{i=1}^k c_i h_{j_0 j_0}(i) \neq 0$. Then the covariance in (20) is not zero if F in Part (d) of the lemma is chosen to correspond to independent errors with $E_F(e_{j_0}^3) \neq 0$ and $E_F(e_j^3) = 0$ for $j \neq j_0$. Again Part (b) of the lemma can then be applied. In case $H_i = H$ for all $i = 1, \dots, k$, these conditions further reduce to $d_{j_0} h_{j_0 j_0} \neq 0$ and $\sum_{i=1}^k c_i \neq 0$.

(iii) Part (c) of Lemma A.1 allows for further examples to be generated, where now the errors need not be independently distributed under the relevant F .

One certainly could set out to characterize those design matrices X for which a counterexample to Assertion 1 can be constructed with the help of Lemma A.1. We do not pursue this here. In particular, we have not investigated whether for *any* $n \times k$ design matrix X with $k < n$ one can construct an estimator $\hat{\beta}_\alpha$ as in the lemma that satisfies (18) for some $c \in \mathbb{R}^k$ and for some $F \in \mathbf{F}_2^0$.

References

- GNOT, S., KNAUTZ, G., 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. *Ann. Math. Statist.*, **17** 34–43.
- HANSEN, B. E. (2021a). *Econometrics*. Princeton University Press, forthcoming. Version August 18, 2021.
- HANSEN, B. E. (2021b). A modern Gauss-Markov theorem, December 2021. Forthcoming in *Econometrica*.
- KAGAN, A. M., LINNIK, Y. V. and RAO, C. R. (1973). *Characterization problems in mathematical statistics*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York-London-Sydney.
- KAGAN, A. M. and SALAEVSKII, O. (1969). The admissibility of least-squares estimates is an exclusive property of the normal law. *Mat. Zametki*, **6** 81–89.
- KOOPMANN, R. (1982). *Parameterschätzung bei a priori Information*. Vandenhoeck & Ruprecht, Göttingen.
- LEHMANN, E. L. and CASELLA, G. (1988). *Theory of Point Estimation*. 2nd ed. Springer-Verlag.
- ROSENDAL, C. (2009). Automatic continuity of group homomorphisms. *The Bulletin of Symbolic Logic*, **15** 184–214.