A Modern Gauss-Markov Theorem? Really?

Pötscher, Benedikt M. and Preinerstorfer, David

University of Vienna, University of St.Gallen

March 2022

A Modern Gauss-Markov Theorem? Really?*

Benedikt M. Pötscher and David Preinerstorfer

Department of Statistics, University of Vienna

SEPS-SEW, University of St. Gallen

First version: February 2022

Second version: March 2022

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 forthcoming in Econometrica) either coincide with the classical theorems just mentioned, or contain extra assumptions that are alien to the Gauss-Markov or Aitken Theorem.

1 Introduction

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, into which he has reintroduced linearity through the backdoor, or contain extra assumptions alien to the Gauss-Markov or Aitken Theorem.

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 form 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 (2021c)).

Hansen (2021a) is the version accepted by Econometrica and – at the time of writing – is available on Econometrica's webpage of forthcoming papers. Hansen (2021b) is an updated version that corrects an incorrect result in Hansen (2021a) (but otherwise is identical to the latter paper), and is available from Hansen's webpage. Hansen (2022), also available at Hansen's webpage, refers to the (corrected) page proofs of the article, which contain several nontrivial changes introduced into the paper at the proof-reading stage. Because Hansen (2021a,b) have been widely circulated and discussed, and because Hansen (2022) will appear in print in Econometrica, there is a need to discuss all three versions. We shall start by first discussing Hansen (2021a,b) in Sections 3 and 4, a discussion that has considerable bearings also on Hansen (2022). We then move on to discuss the changes introduced at the proof-reading stage (Hansen (2022)) and their ramifications in Section 5. Section 6 discusses the situation when one restricts attention to independent identically distributes errors.

After the first version of this paper had been circulated, Stephen Portnoy sent us a paper of his (Portnoy (2022)) that has a result similar to our Theorem 3.4 with a different proof.

2   The Framework

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

$$Y = X\beta + e \tag{1}$$

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

$$Ee = 0 \tag{2}$$

and

$$Eee' = \sigma^2 \Sigma, \tag{3}$$

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

---

[1] We make the assumption $k < n$ in order to use exactly the same framework as in Hansen's papers.

2

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 (2021a,b) 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 (2021a,b))* Hansen (2021a,b) also define 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 (2021a,b) (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 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

$$\Sigma = I_n. \tag{4}$$

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, were 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*

---

[2]Note 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.

$E_F\hat{\beta} = \beta(F)$ 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 (2021a,b)).[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*

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

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., 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:

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

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{OLS}) \tag{6}$$

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

---

[3] As formulated in Hansen (2021a,b), 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.

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

Before discussing Assertions 1 and 2, we need to make a remark on the interpretation of inequalities like (6).

Remark 3.3. (i) In Theorems 3.1 and 3.2 the objects $Var_F(\hat{\beta})$ as well as $Var_F(\hat{\beta}_{OLS})$ are well-defined as real matrices because all estimators considered are linear, and hence $E_F(\| \hat{\beta} \|^2) < \infty$, $E_F(\| \hat{\beta}_{OLS} \|^2) < \infty$ holds for every $F \in \mathbf{F}_2^0$ where $\| . \|$ denotes the Euclidean norm. In contrast, in Assertions 1 and 2 estimators $\hat{\beta}$ with $E_F(\| \hat{\beta} \|^2) = \infty$ for some $F \in \mathbf{F}_2^0$ are permissible. [Note that $E_F(\| \hat{\beta} \|^2) = \infty$ for some $F \in \mathbf{F}_2^0$ and $E_F(\| \hat{\beta} \|^2) < \infty$ for some other $F \in \mathbf{F}_2^0$ may occur.] This necessitates some discussion how Assertion 1 and 2 are then to be read. In the previous version of this paper we unfortunately had glossed over this issue, but an explicit discussion is warranted. For the subsequent discussion note that in both assertions $E_F(\hat{\beta})$ is well-defined and finite for every $F \in \mathbf{F}_2^0$ as a consequence of the respective unbiasedness assumption (and because $\mathbf{F}_2^0 \subseteq \mathbf{F}_2$).

(ii) In the *scalar* case (i.e., $k = 1$), there is no problem as the object $Var_F(\hat{\beta})$ is well-defined for every $F \in \mathbf{F}_2^0$ as an element of the *extended* real line, regardless of whether $E_F(\| \hat{\beta} \|^2) < \infty$ or not. Hence, inequality (6) always makes sense in case $k = 1$.

(iii) For general k, in case the estimator $\hat{\beta}$ satisfies $E_F(\| \hat{\beta} \|^2) < \infty$ for a given $F \in \mathbf{F}_2^0$, the object $Var_F(\hat{\beta})$ is well-defined as a real matrix. Note that the inequality (6) can then equivalently be expressed as $Var_F(c'\hat{\beta}) \geq Var_F(c'\hat{\beta}_{OLS})$ for every $c \in \mathbb{R}^k$.

(iv) In the case $k > 1$, the object $Var_F(\hat{\beta})$ is not well-defined if $E_F(\| \hat{\beta} \|^2) = \infty$ ($F \in \mathbf{F}_2^0$), and hence it is not immediately clear how (6) should then be understood. However, the inequalities $Var_F(c'\hat{\beta}) \geq Var_F(c'\hat{\beta}_{OLS})$ for every $c \in \mathbb{R}^k$ still make sense in view of (ii) above. We hence may and will interpret (6) (with $F \in \mathbf{F}_2^0$) as a symbolic shorthand notation for $Var_F(c'\hat{\beta}) \geq Var_F(c'\hat{\beta}_{OLS})$ for every $c \in \mathbb{R}^k$ (which works both in the case $E_F(\| \hat{\beta} \|^2) < \infty$ and $E_F(\| \hat{\beta} \|^2) = \infty$). We have chosen to write inequality (6) as given (abusing notation), rather than the more conventional and more precise $Var_F(c'\hat{\beta}) \geq Var_F(c'\hat{\beta}_{OLS})$ for every $c \in \mathbb{R}^k$, in order for our discussion to be easily comparable with the presentation in Hansen's papers; his papers are silent on this issue. The same convention applies mutatis mutandis to similar statements such as, e.g., Assertions 3 and 4, etc.

(v) The above discussion would become moot, if one would introduce the *extra* assumption $E_F(\| \hat{\beta} \|^2) < \infty$ for every $F \in \mathbf{F}_2^0$ into Assertions 1 and 2. However, such an additional assumption, which has little justification, would (potentially) narrow down the class of estimators competing with $\hat{\beta}_{OLS}$. As we shall see later on, such an extra assumption actually would have no effect on Assertion 2 at all (and thus on the corresponding theorems in Hansen's papers) in view of our Theorem 3.4. The effect it has on Assertion 1 (and some other results) is discussed

5

in Appendix B.

We now turn to discussing Assertions 1 and 2. Not unexpectedly, Assertion 1 is *incorrect* in general.[4] This is known. For the benefit of the reader we provide some counterexamples and attending discussion in Appendix A. 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 (2021a,b) (his 'modern Gauss-Markov Theorem').[5] Obvious questions now are (i) whether Assertion 2 (i.e., Theorem 5 in Hansen (2021a,b)) is correct, and (ii) if so, what is the reason for Assertion 2 to be correct while Assertion 1 is incorrect in general 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 (2021a,b)) 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. Nevertheless, Hansen (2021a,b) (and also Hansen (2022)) are quiet on the use of this stricter unbiasedness condition.

Having understood what distinguishes Assertion 2 (i.e., Theorem 5 in Hansen (2021a,b)) 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 (2021a,b)) 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., of Theorem 5 in Hansen (2021a,b)) are *linear* estimators! *In other words, 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!* [Note that the word 'linear' does not appear in the *formulation* of Theorem 5 in Hansen (2021a,b), but that linearity of the estimators is introduced indirectly through a backdoor provided by the stricter unbiasedness condition.] While Theorem 5 in Hansen (2021a,b) thus turns out to be correct, it is certainly not new![6] 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 Appendix A shows that this theorem is false. What has been said so far also serves as a reminder that one has to be careful with statements such as "best unbiased equals best linear unbiased".

---

[4] I.e., there exist design matrices X such that the assertion is false.

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

[6] We have not checked whether the *proofs* in Hansen (2021a,b) are correct or not.

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 (2021a,b)) as a consequence of the subsequent theorem.

An upshot of the preceding discussion is that – despite a plea to the contrary in Hansen (2021a,b) – 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 (2021c) is incorrect in general (as it coincides with Assertion 1).

We now provide the theorem alluded to above. After the first version of this paper had been circulated, we learned about Portnoy (2022), which establishes a closely related result using different arguments than the ones we use.

Theorem 3.4. *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).*[7]

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.[8]

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

$$E_F\hat{\beta} = \beta(F) \quad \text{for every} \quad F \in \mathbf{F}_2(\Sigma), \tag{7}$$

for *every* symmetric and positive definite Σ of dimension $n \times n$; specializing to the case $\Sigma = I_n$, we, in particular, obtain[9]

$$E_F\hat{\beta} = \beta(F) \quad \text{for every} \quad F \in \mathbf{F}_2(I_n). \tag{8}$$

Condition (8), together with Theorem 4.3 in Koopmann (1982) (see also Theorem 2.1 in Gnot et al. (1992)[10]), implies that $\hat{\beta}$ is of the form

$$\hat{\beta} = A^0Y + (Y'H_1^0Y, \ldots, Y'H_k^0Y)', \tag{9}$$

where A^0 satisfies $A^0X = I_n$ and H_i^0 are matrices satisfying $\text{tr}(H_i^0) = 0$ and $X'H_i^0X = 0$ for $i = 1, \ldots, k$. It is easy to see that we may without loss of generality assume that the matrices

---

[7]By unbiasedness, such an A must then also satisfy $AX = I_n$.

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

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

[10]Note that X^- in that reference runs through all possible g-inverses of X.

H_i^0 are symmetric (otherwise replace H_i^0 by $(H_i^0 + H_i^{0\prime})/2$). Inserting (9) into (7) yields

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

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

$$(\text{tr}(H_1^0 \Sigma), \ldots, \text{tr}(H_k^0 \Sigma))' = 0 \ \text{ for every symmetric and positive definite } \Sigma. \qquad (10)$$

For every $j = 1, \ldots, n$, 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 $m \to \infty$, where $e_j(n)$ denotes the j-th standard basis vector in $\mathbb{R}^n$ (such sequences obviously exist). Plugging this sequence into (10), letting m go to infinity, and exploiting properties of the trace-operator, we obtain

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

In other words, all the diagonal elements of H_i^0 are zero for every $i = 1, \ldots, k$. Next, for every $j, l = 1, \ldots, n$, $j \neq l$, 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 $m \to \infty$ (such sequences obviously exist). Then exactly the same argument as before delivers

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

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 $i = 1, \ldots, k$. This completes the proof.[11] ∎

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 of our Theorem 3.4 which does not rely on any result in Koopmann (1982).[12]

A direct proof: For every $m \in \mathbb{N}$, every $V = (v_1, \ldots, v_m) \in \mathbb{R}^{n \times m}$ and $\alpha \in (0,1)^m$ such

---

[11] A 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 (7) is of the form $AY + (Y' H_1 Y, \ldots, Y' H_k Y)'$ where $AX = I_n$, the H_i's satisfy $\text{tr}(H_i \Sigma) = 0$, and $X' H_i X = 0$ for $i = 1, \ldots, n$. 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) in principle could depend on Σ. The following argument shows that this is, however, not the case (after symmetrization of the H_i's): If $\hat{\beta}$ had two distinct linear-quadratic representations with symmetric H_i's, 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 (10) in the main text.

[12] Alternatively, 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 our Theorem 3.4.

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 $V \operatorname{diag}(\alpha)V' - (V\alpha)(V\alpha)'$. Denote the expectation operator w.r.t. $\mu_{V,\alpha}$ by $E_{V,\alpha}$. Note that in case $V\alpha = 0$ and $\operatorname{rank}(V) = n$ 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

$$V\alpha = 0 \text{ and } \operatorname{rank}(V) = n \text{ implies } 0 = E_{V,\alpha}(\hat{\beta}) = \sum_{i=1}^{m} \alpha_i \hat{\beta}(v_i). \tag{11}$$

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

$$\mu_{V_2,\alpha^{(2)}} = \frac{n}{n+1} \mu_{V_1,\alpha^{(1)}} + \frac{1}{2(n+1)}(\delta_z + \delta_{-z}). \tag{12}$$

Applying (11) 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

$$\hat{\beta}(-z) = -\hat{\beta}(z) \text{ for every } z \in \mathbb{R}^n, \tag{13}$$

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

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

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

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

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

Then, we obtain $V\alpha = 0$ and $\operatorname{rank}(V) = n$. Using (11) and (13) 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 (13) is equivalent to

$$\hat{\beta}(y) + \hat{\beta}(z) = \sum_{i=1}^{n} \hat{\beta}((y_i + z_i)e_i(n)). \tag{14}$$

9

Using (14) 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

$$\hat{\beta}(y) + \hat{\beta}(z) = \hat{\beta}(y+z) \quad \text{for every } y \text{ and } z \text{ in } \mathbb{R}^n. \tag{15}$$

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 (15) (which contains (13) 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 $\hat{\beta}$

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

This concludes the proof. ∎

Remark 3.5. Inspection of the direct proof above shows that it does not make use of the full force of the unbiasedness condition ($E_F\hat{\beta} = \beta(F)$ for every $F \in \mathbf{F}_2$), but only exploits unbiasedness for certain strategically chosen discrete distributions F, each with finite support and satisfying $\beta(F) = 0$.

Remark 3.6. *(Ambiguity in the definition in Hansen (2021a,b) 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 in Hansen (2021a,b) 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 (2021a,b) 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 (2021a,b).

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 (2021a,b), his rendition of the Aitken Theorem (for linear estimators). Note that the subsequent theorem is obviously equivalent to Theorem 3 in Hansen (2021a,b) 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 (16) below is understood as $(\sigma^2\Sigma)(F)$, as is necessary anyways for Theorem 3 in Hansen (2021a,b) 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*

$$Var_F(\hat{\beta}) \succeq Var_F(\hat{\beta}_{GLS}) \tag{16}$$

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 (2021a,b) (to which we shall refer as his 'modern Aitken Theorem'); this is seen in the same way as the equivalence of Theorem 4.2 above with Theorem 3 in Hansen (2021a,b). Assertion 4 is indeed correct, but again not new, as the class of estimators figuring in Assertion 4 consists only of linear estimators as a consequence of Theorem 3.4 above.[13] Furthermore, a comment like Remark 3.6 also applies here. We conclude this section by noting that the rendition of Aitken's Theorem in the text-book Hansen (2021c) (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 The Results in Hansen (2022)

In Hansen (2022) the same model given by (1), (2), and (3) as in Hansen (2021a,b) is considered and $\mathbf{F}_2$ is defined in the same manner.[14] A set $\mathbf{F}_2^*$ representing the subset of $\mathbf{F}_2$ corresponding to *independent* errors $e_1, \ldots, e_n$ is also defined; here e_i denotes the i-th component of the error vector e. Furthermore, the subset of $\mathbf{F}_2^*$ corresponding to *independent homoskedastic* errors is denoted by $\mathbf{F}_2^0$. It should be noted that this set is *not* the same as the set $\mathbf{F}_2^0$ in Hansen (2021a,b). To avoid any confusion we shall in the following write $\mathbf{F}_2^{0,new}$ for the set denoted by $\mathbf{F}_2^0$ in Hansen (2022).

We start with a discussion of the treatment of Aitken's Theorem in Hansen (2022). Hansen first gives a rendition of the classical Aitken Theorem (Theorem 3 in Hansen (2022)) which is identical to Theorem 3 in Hansen (2021a,b), and thus to Theorem 4.2 in the preceding section. He proceeds to provide his 'modern Aitken Theorem' (Theorem 4 in Hansen (2022)), which is identical to the corresponding Theorem 4 in Hansen (2021a,b) and which in turn is equivalent to Assertion 4 as just discussed in Section 4 above.[15] Consequently, the discussion in Section 4 above applies. In particular, the estimators figuring in the 'modern Aitken Theorem' in Hansen (2022) are all automatically linear by our Theorem 3.4, and hence the 'modern Aitken Theorem' in Hansen (2022) is not new, but reduces to the classical Aitken Theorem.

Hansen then goes on to provide a further result (Theorem 5 in Hansen (2022)) which can equivalently be stated as the following assertion (the equivalence is seen in the same way as the

---

[13]Adding the extra condition $E_F(\| \hat{\beta} \|^2) < \infty$ for every $F \in \mathbf{F}_2(\Sigma)$ has no effect on Assertion 4 in view of our Theorem 3.4. The effect this extra condition has on Assertion 3 is discussed in Appendix B.

[14]The assumption in Hansen (2021a,b) that Σ is known and positive definite and that σ^2 is positive has been dropped in Hansen (2022). Nevertheless positive definiteness of Σ as well as $\sigma^2 > 0$ are frequently used in Hansen (2022) (e.g., inverses of Σ are taken; the proof of Theorem 4 makes use of both properties). We hence will continue to assume $\sigma^2 > 0$ and positive definiteness of Σ in our discussion. We furthermore note that the ambiguity in the definition of $\mathbf{F}_2$ in Hansen (2021a,b) is now being avoided in Hansen (2022) as Σ is no longer assumed to be known. Of course, then σ^2 and Σ are no longer identifiable.

[15]The same caveat regarding the formulation of Hansen's theorems as discussed in Section 4 applies here.

equivalence between Theorem 3 in Hansen (2021a,b) and Theorem 4.2 in Section 4 above).[16] For Σ a diagonal $n \times n$ matrix with positive diagonal elements, define $\mathbf{F}_2^*(\Sigma) = \{F \in \mathbf{F}_2^* : Var_F(e) \propto \Sigma\}$, where $\propto$ denotes proportionality. Of course, then $\mathbf{F}_2^* = \bigcup\{\mathbf{F}_2^*(\Sigma) : \Sigma$ diagonal with positive diagonal elements$\}$ and $\mathbf{F}_2^*(I_n) = \mathbf{F}_2^{0,new}$. Recall that $\hat{\beta}_{GLS} = \hat{\beta}_{GLS}(\Sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$.

Assertion 5: Let Σ be an arbitrary known *diagonal* $n \times n$ matrix with positive diagonal elements. 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 $F \in \mathbf{F}_2^*(\Sigma)$.

This assertion is very much different from an Aitken Theorem:[17] (i) The errors in the model corresponding to $F \in \mathbf{F}_2^*(\Sigma)$ (i.e., distributions F for which the variance inequality has to hold) need to be *independent*, an assumption alien to Aitken's Theorem that allows for general Σ and also relies only on first and second moment assumptions (as opposed to an independence assumption), and (ii) the unbiasedness assumption is – like in results discussed earlier – required to hold under the *wider* class of distributions $\mathbf{F}_2^*$, and not only under the distributions F describing the data generating mechanism (i.e., $F \in \mathbf{F}_2^*(\Sigma)$). While in the context of Assertions 2 and 4 no nonlinear unbiased estimator exists, in the context of Assertion 5 nonlinear unbiased estimators indeed exist (at least for some matrices X), cf. Remark 5.1 below.

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). 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 $F \in \mathbf{F}_2^*$ is identical to requiring unbiasedness for all $F \in \mathbf{F}_2^{0,new}$ (or for all $F \in \mathbf{F}_2^0 = \mathbf{F}_2(I_n)$ for that matter); cf. the discussion following Theorem 3.2 in Section 3.] The 'modern Gauss-Markov Theorem' (Theorem 6 in Hansen (2022)) is now the special case of Assertion 5 for $\Sigma = I_n$; note that this result is just Theorem 1 in Hansen (2022) with the linearity requirement dropped.[18] For the same reasons as discussed in the preceding paragraph in connection with Assertion 5, this result can not legitimately be called a (modern) Gauss-Markov Theorem. Finally, Theorem 7 in Hansen (2022) just specializes Theorem 6 in the same reference to the location model, hence the same remarks apply.

---

[16] The same caveat regarding the formulation of Hansen's theorems discussed in Section 4 applies also to Theorem 5 in Hansen (2022).

[17] This assertion (equivalently, Theorem 5 in Hansen (2022)) seems to be correct. However, we have not checked the correctness of the proofs in Hansen (2022) in any detail.

[18] The 'modern Gauss-Markov Theorem' as given in Theorem 5 of Hansen (2021a,b) is no longer presented, but of course is an immediate consequence of Theorem 4 in Hansen (2022). Also Theorem 6 of Hansen (2021a,b) is no longer given.

To sum up, Theorems 4-7 in Hansen (2022) are either an intransparent restatement of the classical Aitken Theorem introducing linearity of the estimators through the backdoor (Theorem 4 in Hansen (2022)), or are results modelled on the Gauss-Markov or Aitken Theorem but with *substantial extra* conditions added (Theorems 5-7 in Hansen (2022)). The significance and scope of the latter results is unclear for the reasons discussed before and by no means justify the advertisements made in Hansen (2022). In particular, the claim made in the abstract and repeated at the end of Section 5 of Hansen (2022), that his theorems would show that the label "linear estimator" can be dropped from the pedagogy of the Gauss-Markov Theorem, is without any base. We thus repeat our warning against dropping the linearity assumption from the Gauss-Markov or Aitken Theorem.

Remark 5.1. In the discussion following Theorem 5 in Hansen (2022), the author gives an example of a nonlinear estimator that is unbiased under all $F \in \mathbf{F}_2^*$. The object $\tilde{\beta}$ given there, however, is not well-defined as it is the sum of two components that are of different dimension (unless $k = 1$). This can be rectified by redefining $\tilde{\beta}$ as $\hat{\beta}_{OLS} + Y_i(Y_j - X_j'\hat{\beta}_{-i})a$ $(i \neq j)$ for any chosen $k \times 1$ vector $a \neq 0$ (here Y_j and X_j' denote the j-th row of Y and X, respectively).[19] This object is indeed unbiased under all $F \in \mathbf{F}_2^*$. The claim in Hansen (2022) that this is a *nonlinear* estimator, however, is not generally true for any design matrix X. For example, if $n = k + 1$, then any leave-one-out residual is zero, and hence $\tilde{\beta} = \hat{\beta}_{OLS}$ is linear (for any choice of i and j). Another example where $\tilde{\beta} = \hat{\beta}_{OLS}$ is when $k = 1$, the regressor is the first standard basis vector, $i \neq 1$, and $j = 1$. Fortunately, there are examples of design matrices for which $\tilde{\beta}$ is indeed truly nonlinear.

6   Independent Identically Distributed Errors

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

$$e_1, \ldots, e_n \quad \text{are i.i.d.} \tag{17}$$

to the model. Let $\mathbf{F}_2^{iid}$ be the subset of $\mathbf{F}_2^0$ corresponding to distributions F that result from (1), (2), (3), and (17). In particular, we ask what is the status of the following assertion which is analogous to Assertion 1.

Assertion 6: 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 $F \in \mathbf{F}_2^{iid}$.

---

[19] There is an implicit assumption here, namely that the design matrix continues to have full column rank even after the i-th row is deleted.

Note that Assertion 6 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 6 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 1. Hence, the truth-status of Assertion 1 does not inform us about the corresponding status of Assertion 6.

Fortunately, Example A.2 in Appendix A comes to the rescue and shows that Assertion 6 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 6 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 Appendix A.]

For results in the location case pertaining to classes of absolutely continuous distributions (without or with symmetry restrictions) see Example 4.2 in Section 2.4 of Lehmann and Casella (1988) and the discussion following this example.

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 (and has finite second moments). 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). For more results in that direction see Sections 7.4-7.9 in the same reference.

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), Knautz (1993, 1999), 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

$$\hat{\beta}_\alpha = \hat{\beta}_{OLS} + \alpha(Y'H_1Y, \ldots, Y'H_kY)' \tag{18}$$

where the H_i's are symmetric $n \times n$ matrices and α is a real number. Suppose $\mathrm{tr}(H_i) = 0$ and $X'H_iX = 0$ for $i = 1, \ldots, k$. 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 $Cov_F(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \ldots, Y'H_kY)') \neq 0$ for some $c \in \mathbb{R}^k$ and for some $F \in \mathbf{F}_2^0$ with finite fourth moments, then there exists an $\alpha \in \mathbb{R}$ such that

$$Var_F(c'\hat{\beta}_\alpha) < Var_F(c'\hat{\beta}_{OLS}); \tag{19}$$

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 $F \in \mathbf{F}_2^0$ (with finite fourth moments) under which $\beta(F) = 0$ we have

$$Cov_F\left(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \ldots, Y'H_kY)'\right) = \sum_{j=1}^n \sum_{l=1}^n \sum_{m=1}^n d_j \left(\sum_{i=1}^k c_i h_{lm}(i)\right) E_F(e_j e_l e_m), \tag{20}$$

where $d = (d_1, \ldots, d_n)' = X(X'X)^{-1}c$ 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 $F \in \mathbf{F}_2^0$ (with finite fourth moments) under which (i) $\beta(F) = 0$ and under which (ii) the coordinates of Y are independent (equivalently, the errors e_i are independent)

$$Cov_F\left(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \ldots, Y'H_kY)'\right) = \sum_{j=1}^n d_j \left(\sum_{i=1}^k c_i h_{jj}(i)\right) E_F(e_j^3). \tag{21}$$

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

$$\begin{aligned} Var_F(c'\hat{\beta}_\alpha) &= Var_F(c'\hat{\beta}_{OLS}) + 2\alpha Cov_F\left(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \ldots, Y'H_kY)'\right) \\ &\quad + \alpha^2 Var_F(c'(Y'H_1Y, \ldots, Y'H_kY)'), \end{aligned} \tag{22}$$

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 $2Cov_F\left(c'\hat{\beta}_{OLS}, c'(Y'H_1Y, \ldots, Y'H_kY)'\right)$. Note that all terms in (22) are well-defined and finite because of our fourth moment assumption. Hence, whenever this covariance is non-zero, we may choose $\alpha \neq 0$ small enough such that (19) 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 $X = (1, \ldots, 1)'$. 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 (21)

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

for every $F \in \mathbf{F}_2^0$ (with finite fourth moments) 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.[20] As a consequence,

[20]E.g., choose $e_2, \ldots, e_n$ i.i.d. $N(0, \sigma^2)$ and e_1 independent from $e_2, \ldots, e_n$ with mean zero, variance σ^2, third

$\hat{\beta}_{OLS}$ 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., $X = (1, \ldots, 1)'$) 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 $F \in \mathbf{F}_2^0$ (with finite fourth moments) under which the errors e_i are i.i.d., since then $Cov_F(\hat{\beta}_{OLS}, Y'H_1Y) = 0$ as is easily seen. [This is in line with the result of Halmos (1946) discussed in Section 6.] 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.

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 (21) we find

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

under any $F \in \mathbf{F}_2^0$ (with finite fourth moments) 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 (21) 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 $i = 1, \ldots, k$, 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

moment not equal to zero, and finite fourth moment.

(21) 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 $i = 1, \ldots, k$, 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 (19) for some $c \in \mathbb{R}^k$ and for some $F \in \mathbf{F}_2^0$.

B Appendix: Finite Second Moments of Estimators

We start by discussing the consequences of introducing the requirement $E_F(\| \hat{\beta} \|^2) < \infty$ for every $F \in \mathbf{F}_2^0$ into Assertions 1 and 2. First, note that nothing changes for Assertion 2, since the estimators $\hat{\beta}$ allowed in Assertion 2 are all linear as a consequence of our Theorem 3.4. Turning to Assertion 1, we first give the following proposition.

Proposition B.1. *Consider estimators of the form*

$$\hat{\beta} = AY + (Y'H_1Y, \ldots, Y'H_kY)'$$

with $AX = I_n$ and the $n \times n$ matrices H_j satisfying $\mathrm{tr}(H_j) = 0$ and $X'H_jX = 0$ for $j = 1, \ldots, k$. Suppose $E_F(\| \hat{\beta} \|^2) < \infty$ for every $F \in \mathbf{F}_2^0$ holds. Then $\hat{\beta}$ is a linear estimator.

Proof: First, observe that $E_F(\hat{\beta}) = \beta(F)$ for all $F \in \mathbf{F}_2^0$ holds. Second, we may assume the matrices H_j to be symmetric (if necessary we replace H_j by $(H_j + H_j')/2$). Since AY obviously has finite second moments under every $F \in \mathbf{F}_2^0$, the finite second moment assumption on $\hat{\beta}$ implies that $E_F((Y'H_jY)^2) < \infty$ has to hold for every $F \in \mathbf{F}_2^0$ and every $j = 1, \ldots, k$. Let now j $(j = 1, \ldots, k)$ be arbitrary, but fixed. By symmetry of H_j, there exists an orthogonal matrix U such that $UH_jU' = \Lambda_j$ where Λ_j is a diagonal matrix. Set $Z = UY$ and let λ_{ij} denote the elements on the diagonal of Λ_j. Then

$$(Y'H_jY)^2 = (Z'\Lambda_jZ)^2 = \left(\sum_{i=1}^n \lambda_{ij} z_i^2 \right)^2 = \sum_{i=1}^n \lambda_{ij}^2 z_i^4 + \sum_{i,l=1, i \neq l}^n \lambda_{ij} \lambda_{lj} z_i^2 z_l^2. \tag{23}$$

Let $F \in \mathbf{F}_2^0$ be such that $\beta(F) = 0$ and such that the elements of Z are i.i.d. with mean zero, finite variance, and infinite fourth moment. Such an F exists: Start with a distribution on Z with the required properties just listed and work backwards, defining $Y = U'Z$. Then clearly the implied F has $\beta(F) = 0$ and belongs to $\mathbf{F}_2^0$. Since the coordinates of Z are independent and have a finite second moment, the term $\sum_{i,l=1, i \neq l}^n \lambda_{ij} \lambda_{lj} z_i^2 z_l^2$ has finite expectation under F.

18

Since the l.h.s. of (23) has finite expectation under F under our assumptions as noted before, $E_F(\sum_{i=1}^{n} \lambda_{ij}^2 z_i^4)$ has to be finite. Since $E_F(z_i^4) = \infty$ for every $i = 1, \ldots, n$, we must have $\lambda_{ij} = 0$ for every i and the given j. This shows that $H_j = 0$. Since j was arbitrary, this holds for every j, and thus $\hat{\beta} = AY$ is linear. ∎

The significance of the preceding proposition is the following. Suppose Theorem 4.3 in Koopmann (1982) is indeed correct (recall that no complete proof is given in Koopmann (1982)). Then introducing the *extra* condition of finite second moments for the estimators $\hat{\beta}$ (under every $F \in \mathbf{F}_2^0$) into Assertion 1 reduces the class of competitors to the class of linear unbiased estimators. The resulting *version* of Assertion 1 is thus true and coincides with the classical Gauss-Markov Theorem. [Recall that Assertion 1 is false in general.] Again, linearity is reintroduced indirectly by adding the before mentioned extra condition. In case Theorem 4.3 in Koopmann (1982) is incorrect, then other estimators than quadratic ones might figure in this version of Assertion 1 and it is unclear whether this version of Assertion 1 is true or not, and if it is true what its scope is.

Concerning Assertion 3, an analogous version of the preceding proposition can be given and a similar discussion applies. Note that Assertion 4 and the corresponding theorems in Hansen's papers are not affected by introducing an extra finite second moment assumption on the estimators, because all estimators in Assertion 4 are automatically linear in view of our Theorem 3.4.

Adding an extra finite second moment assumption on the estimators to Assertion 5 leads to a version that, a fortiori, is not an 'Aitken Theorem', for the same reasons as discussed in Section 5. [In case Assertion 5 is correct, this version is a fortiori also correct.] Whether or not Assertion 6, which is false in general, becomes a true statement after adding an extra finite second moment condition on the estimators, we have not investigated.

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). A modern Gauss-Markov theorem, September 2021. Version accepted for publication in Econometrica. URL `https://www.econometricsociety.org/system/files/19255-3.pdf`.

HANSEN, B. E. (2021b). A modern Gauss-Markov theorem, December 2021. Update of September 2021 version accepted for publication in Econometrica. URL `https://www.ssc.wisc.edu/~bhansen/papers/gauss.pdf`.

HANSEN, B. E. (2021c). *Econometrics*. Princeton University Press, forthcoming. Version August 18, 2021.

HANSEN, B. E. (2022). A modern Gauss-Markov theorem. Version forthcoming in Econometrica. URL https://www.ssc.wisc.edu/~bhansen/papers/gauss2.pdf.

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.

KNAUTZ, H. (1993). *Nichtlineare Schätzung des Parametervektors im linearen Regressionsmodell*, vol. 133 of *Mathematical Systems in Economics*. Verlag Anton Hain, Frankfurt am Main.

KNAUTZ, H. (1999). Nonlinear unbiased estimation in the linear regression model with nonnormal disturbances. *J. Statist. Plann. Inference*, **81** 293–309.

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.

PORTNOY, S. (2022). Linearity of unbiased linear model estimators, February 26, 2022.

ROSENDAL, C. (2009). Automatic continuity of group homomorphisms. *The Bulletin of Symbolic Logic*, **15** 184–214.