Econometrica, Vol. 92, No. 3 (May, 2024), 925–928

REPLY TO: Comment on *"A Modern Gauss–Markov Theorem"*

BRUCE E. HANSEN
Department of Economics, University of Wisconsin

This note makes a brief response to Portnoy (2022) and Pötscher and Preinerstorfer (2024), and discusses what instructors should teach about best unbiased estimation.

1. JOINT DEPENDENCE

HANSEN (2022A) ESTABLISHED A SET of finite-sample efficiency lower bounds for the linear regression model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$ with fixed regressors and finite variance matrix $\text{var}[\boldsymbol{e}] = \sigma^2\boldsymbol{\Sigma}$. These results cover the cases of joint dependence with unrestricted $\boldsymbol{\Sigma}$, and independent sampling with diagonal $\boldsymbol{\Sigma}$.

One of these results (Theorem 4) demonstrates, in the context of joint dependence with unrestricted $\boldsymbol{\Sigma}$, that an unbiased estimator of β cannot have a lower variance than $\sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$. As this is the variance of the generalized least squares (GLS) estimator, it follows that the latter is the best unbiased estimator (BUE) of β. No explicit restriction to linear estimators is necessary.

In a pair of insightful papers, Portnoy (2022) and Pötscher and Preinerstorfer (2024) showed that in the specific context of Theorem 4, all unbiased estimators of β are linear estimators. Since the lowest variance among unbiased linear estimators is $\sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$, this can be viewed as an alternative proof that the GLS estimator is the BUE. The fact, however, that an unbiased estimator must be linear severely limits the relevance of Theorem 4.

2. INDEPENDENT SAMPLING

Another set of results (Theorems 5–7) in Hansen (2022a) examine the case of independent sampling. Neither Portnoy (2022) nor Pötscher and Preinerstorfer (2024) examined this case. For clarity, it is useful to review and explain the main result.

In this setting, the variables Y_i are mutually independent across i and satisfy the linear regression $Y_i = X_i'\beta + e_i$ with X_i fixed, $\mathbb{E}[e_i] = 0$, and $\mathbb{E}[e_i^2] = \sigma_i^2$. The variances satisfy $0 < \sigma_i^2 < \infty$ but are otherwise unrestricted. Let F denote the joint distribution of $(Y_1, \ldots, Y_n)$. The class of joint distributions F satisfying these conditions is denoted as $\mathbf{F}_2^*$. This is the class of linear regression models with possibly heteroskedastic variances. The class $\mathbf{F}_2^*$ fixes the regressors X_i, but includes all possible regression coefficients β, error variances σ_i^2, and error distributions. The variables Y_i are mutually independent and satisfy a linear regression $Y_i = X_i'\beta + e_i$, but otherwise their distributions are unrestricted.

Now consider unbiased estimation of the regression coefficient β. An estimator $\widehat{\beta}$ is unbiased in the class $\mathbf{F}_2^*$ if $\mathbb{E}[\widehat{\beta}] = \beta$ for all distributions $F \in \mathbf{F}_2^*$. This means that $\widehat{\beta}$ is unbiased for β whenever the joint distribution satisfies a linear regression model. An example is GLS:

$$\widehat{\beta}_{\text{gls}} = \left(\sum_{i=1}^{n} \sigma_i^{-2} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} \sigma_i^{-2} X_i Y_i\right). \tag{1}$$

Bruce E. Hansen: bruce.hansen@wisc.edu

(The GLS estimator is infeasible in practice, but is a useful theoretical benchmark.) The estimator $\widehat{\beta}_{\text{gls}}$ is called a *linear estimator* as it is a linear function of the dependent variables $(Y_1, \ldots, Y_n)$.

Given the Portnoy–Pötscher–Preinerstorfer result, it is reasonable to ask if there exist unbiased nonlinear estimators. The answer is yes. For example, take the location model $X_i = 1$ and consider

$$\widehat{\beta}_1 = \widehat{\beta}_{\text{gls}} + Y_1 Y_2 - Y_3 Y_4, \tag{2}$$

which is a nonlinear function of $(Y_1, \ldots, Y_n)$. A simple calculation reveals that $\mathbb{E}[\widehat{\beta}_1] = \beta$ for all $F \in \mathbf{F}_2^*$, so $\widehat{\beta}_1$ is unbiased. Another calculation reveals that $\text{var}[\widehat{\beta}_1] > \text{var}[\widehat{\beta}_{\text{gls}}]$ for any $F \in \mathbf{F}_2^*$. Thus, the addition of nonlinearity increases estimation variance.

Theorem 5 of Hansen (2022a) shows that this holds for all unbiased estimators.

THEOREM 5: *If $\widehat{\beta}$ is unbiased for all $F \in \mathbf{F}_2^*$, then $\text{var}[\widehat{\beta}] \geq \text{var}[\widehat{\beta}_{\text{gls}}]$ for all $F \in \mathbf{F}_2^*$.*

No unbiased estimator has lower variance than GLS, and therefore GLS is the *best (lowest variance) unbiased estimator*. Theorem 5 makes no restriction to linear estimators; there is no restriction other than unbiasedness. However, Theorem 5 is not a strict improvement on the classical Gauss–Markov theorem as the latter only requires uncorrelated samples, while $\mathbf{F}_2^*$ restricts attention to independent samples.

Another sharp result can be obtained in the location model with i.i.d. sampling. Let Y_i be i.i.d. with distribution F, population mean $\mathbb{E}[Y_i] = \beta$, and variance $\text{var}[Y_i] = \sigma^2 < \infty$. Let $\mathbf{F}_2^0$ be the class of distributions F with a finite variance. An estimator $\widehat{\beta}$ is unbiased in the class $\mathbf{F}_2^0$ if $\mathbb{E}[\widehat{\beta}] = \beta$ for all distributions $F \in \mathbf{F}_2^0$. An example is the sample mean $\overline{Y}$. Theorem 11.1 of Hansen (2022b) shows that no unbiased estimator has a lower variance.

THEOREM 11.1: *If $\widehat{\beta}$ is unbiased for all $F \in \mathbf{F}_2^0$, then $\text{var}[\widehat{\beta}] \geq \text{var}[\overline{Y}]$ for all $F \in \mathbf{F}_2^0$.*

Theorem 11.1 shows that the sample mean $\overline{Y}$ is the *best unbiased estimator* of the population mean under i.i.d. sampling. No restriction to linear estimators is necessary. Theorem 11.1 is also a strict improvement over the Cramér–Rao theorem (e.g., Theorem 2 of Hansen (2022a)), as Theorem 11.1 holds for all distributions, while the Cramér–Rao theorem requires normality.

Appendix A of Pötscher and Preinerstorfer (2024) presents a related but distinct example which provides additional insights. Take the location model $X_i = 1$, assume homoskedastic variances $\sigma_i^2 = 1$, and consider the nonlinear estimator

$$\widehat{\beta}_2 = \overline{Y} + \frac{Y_1^2 - Y_2^2}{n}. \tag{3}$$

A simple calculation reveals that under the stated assumptions, $\mathbb{E}[\widehat{\beta}_2] = \beta$, so $\widehat{\beta}_2$ is unbiased in this class. To calculate the variance of $\widehat{\beta}_2$, for simplicity assume $\beta = 0$, e_1 has the Rademacher distribution, and e_2 the Mammen (1993) distribution.[1] A straightforward calculation shows that under these conditions,

$$\text{var}[\widehat{\beta}_2] = \frac{1}{n} - \frac{1}{n^2} < \text{var}[\overline{Y}], \tag{4}$$

[1]Which satisfy $\mathbb{E}[e_1^3] = 0$, $\mathbb{E}[e_1^4] = 1$, $\mathbb{E}[e_2^3] = 1$, and $\mathbb{E}[e_2^4] = 2$.

showing that $\widehat{\beta}_2$ has a lower variance than the sample mean. At first glance, this may appear to contradict Theorem 5 and/or Theorem 11.1, but it does not. First, while $\widehat{\beta}_2$ is unbiased under the assumption of homoskedastic variances, it is biased under heteroskedastic variances. Therefore, it is not unbiased in the class of models $\mathbf{F}_2^*$, and hence falls outside the scope of Theorem 5, so is not a counterexample to Theorem 5. The estimator $\widehat{\beta}_2$ is able to achieve improved efficiency only by sacrificing unbiasedness under heteroskedasticity. Second, the calculation (4) exploits the assumption that the observations Y_1 and Y_2 have different third moments; when they are identically distributed, then $\text{var}[\widehat{\beta}_2] \geq \text{var}[\overline{Y}]$. Consequently, this example falls outside the scope of Theorem 11.1, so is not a counterexample to Theorem 11.1.

Together, the examples (2) and (3) illustrate the powerful role of unbiasedness in Theorems 5 and 11.1, and the role of identical distributions in Theorem 11.1.

3. WHAT SHOULD WE TEACH?

The reason why instruction includes the BLUE and Gauss–Markov theorems is because we want simple justifications for standard estimators. The BLUE and Gauss–Markov theorems are awkward for this purpose because of the unnatural restriction to linear estimators.

This material is typically taught in the context of independent sampling, where Theorems 5 and 11.1 are relevant. However, the phrasing of these theorems as presented in the previous section, while precise, may be overly technical for instruction. Instead, I believe that the following simplified restatements can be constructively used.

First, take estimation of the population mean under i.i.d. sampling, which is typically discussed in introductory classes.

THEOREM 11.1′: *Let Y_i be i.i.d. with a finite variance. Then, the sample mean $\overline{Y}$ is the* ***best unbiased estimator*** *of the population mean $\mathbb{E}[Y]$.*

When taught, it should be explained that "best" refers to "minimum variance," and "unbiased"refers to "unbiased under i.i.d. sampling from any distribution with a finite variance." Theorem 11.1′ can be presented to students to justify why the sample mean is the standard estimator of the population mean. We can replace the BLUE acronym with BUE.

Second, take the case of linear regression with independent but not necessarily identically distributed observations, which is typically discussed in intermediate econometrics classes.

THEOREM 5′: *Let $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ be an independent sample satisfying a linear regression $Y_i = X_i'\beta + e_i$ with $\mathbb{E}[e_i] = 0$, $\mathbb{E}[e_i^2] = \sigma_i^2$, and $0 < \sigma_i^2 < \infty$. Then, the* ***best unbiased estimator*** *of β is GLS (1).*

When taught, it should be explained that "unbiased" refers to "unbiased under independent sampling from any linear regression with possibly heteroskedastic variances." It is important to understand that this unbiased property must hold under any form of heteroskedasticity. Theorem 5′ can be used in instruction to demonstrate why we focus on GLS estimators. Theorem 5′ can also be used to deduce that when the error variances are homoskedastic, then the BUE is ordinary least squares.

A reasonable question is whether or not instructors will want to discuss the proofs of Theorems 5 and 11.1. The most accessible presentation of Theorem 11.1 can be found

in Section 11.6 of Hansen (2022b), and that of Theorem 5 in Sections 4.8–4.9 of Hansen (2022c). While these textbook treatments focus on the independent sampling case, they are still quite advanced. For many levels of instruction, therefore, it may be prudent to skip the proof, assert that the BUE is linear, and then proceed with the conventional derivation of the best unbiased estimator among linear estimators.

REFERENCES

HANSEN, BRUCE E. (2022a): "A Modern Gauss-Markov Theorem," *Econometrica*, 90, 1283–1294. [925,926]

——— (2022b): *Probability and Statistics for Economists*. Princeton University Press. [926,928]

——— (2022c): *Econometrics*. Princeton University Press. [928]

MAMMEN, ENNO (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *Annals of Statistics*, 21, 255–285. [926]

PORTNOY, STEPHEN (2022): "Linearity of Unbiased Linear Model Estimators," *The American Statistician*, 76, 372–375. [925]

PÖTSCHER, BENEDIKT M., AND DAVID PREINERSTORFER (2024): "On 'A Modern Gauss-Markov Theorem' by B. E. Hansen," Econometrica. (forthcoming). [925,926]

Co-editor Guido W. Imbens handled this manuscript.

Manuscript received 17 October, 2023; final version accepted 18 February, 2024; available online 22 February, 2024.