Journal of Statistical Computation and Simulation
Vol. 79, No. 3, March 2009, 299–314

Taylor & Francis
Taylor & Francis Group

Anomalies in the analysis of calibrated data

D.R. Jensen and D.E. Ramirez*

Department of Mathematics, University of Virginia, Charlottesville, VA, USA

(*Received 30 November 2006; final version received 18 October 2007*)

This study examines the effects of calibration errors on model assumptions and data-analytic tools in direct calibration assays. These effects encompass induced dependencies, inflated variances, and heteroscedasticity among the calibrated measurements, whose distributions arise as mixtures. These anomalies adversely affect conventional inferences, including the inconsistency of sample means; the underestimation of measurement variance; and the distributions of sample means, sample variances, and student's t as mixtures. Inferences in comparative experiments remain largely intact, although error mean squares continue to underestimate the measurement variances. These anomalies are masked in practice, as conventional diagnostics cannot discern the irregularities induced through calibration. Case studies illustrate the principal issues.

Keywords: direct calibration; induced dependencies; non-standard distributions; diagnostics; case studies

1991 Mathematics Subject Classifications: Primary: 62J05; Secondary: 62E10

1. Introduction

Calibrated measurements, intrinsic to the sciences and engineering, are inherently subject to errors of calibration. These errors induce dependencies in the violation of a basic tenet in much of applied statistics, namely, that observations should be uncorrelated if not independent. These issues traditionally have been overlooked by both scientists and statisticians, despite a century of emerging methodologies for the analysis of experimental data. Not only are many parametric and non-parametric procedures at risk under such violations, but so also are conventional diagnostics for checking critical features of a model. We return to these subsequently.

To fix ideas, observed responses $\{Z_0, Z_1, \ldots, Z_n\}$ are often adjusted to Z_0 as standard, giving differences $\{Y_i = (Z_i - Z_0);\ 1 \le i \le n\}$ as the objects of interest to the investigator. Moreover, if $\{Z_0, Z_1, \ldots, Z_n\}$ are mutually uncorrelated having variances $\{\mathrm{var}(Z_0) = \sigma_0^2,\ \mathrm{var}(Z_i) = \sigma^2;\ 1 \le i \le n\}$ then $\{Y_1, \ldots, Y_n\}$ are equicorrelated with the parameter $\rho = \sigma_0^2/(\sigma^2 + \sigma_0^2)$ having variances $\{\mathrm{var}(Y_i) = \sigma^2 + \sigma_0^2; 1 \le i \le n\}$ that are inflated in comparison with unadjusted values.

Linearly calibrated instruments are pervasive. Some unintended consequences, to be examined here, include (1) the structure of induced dependencies, heteroscedasticity, and other departures

*Corresponding author. Email: der@virginia.edu

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650701756864
http://www.informaworld.com

from conventional model assumptions; (2) the inflation of measurement variances in comparison with intended values; and (3) effects of calibration on conventional inferences for location, scale, and model diagnostics. We first examine moments, then effects of calibration on the actual measurement distributions themselves. We focus here on *direct calibration assays* to be identified subsequently. An outline follows.

Section 2 gives notation and technical support. Section 3 reconsiders the calibration process with reference to the irregularities induced through calibration errors. Section 4 addresses the impact of these irregularities on conventional inferences, including (1) inferences regarding the mean and variance in a single sample and (2) the analysis of one–way experimental data, including tests on means and variances. The latter remain largely intact, although measurement variances continue to be underestimated. Section 5 re-examines the ability of conventional diagnostics to uncover the violations induced through calibration. Section 6 undertakes a case study to illustrate essential findings. Section 7 offers a brief summary and cautionary note. Some peripheral matters are deferred to an Appendix.

2. Preliminaries

2.1. *Notation*

Designate $\mathbb{R}^n$ as an Euclidean n-space, $\mathbb{R}^n_+$ as its positive orthant, $\mathbb{S}_n$ as the real symmetric $(n \times n)$ matrices, and $\mathbb{S}_n^+$ and $\mathbb{S}_n^0$ as their positive definite and positive semidefinite varieties. Arrays appear in bold type, including the transpose $\boldsymbol{A}'$ and inverse $\boldsymbol{A}^{-1}$ of $\boldsymbol{A}$, the unit vector $\mathbf{1}_n = [1, \ldots, 1]' \in \mathbb{R}^n$, the identity matrix $\boldsymbol{I}_n$, a block-diagonal matrix diag $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$, and $\boldsymbol{B}_n = (\boldsymbol{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')$. Following Loewner [1], matrices $(\boldsymbol{A}, \boldsymbol{B})$ in $\mathbb{S}_n$ are said to be ordered as $\boldsymbol{A} \succeq_L \boldsymbol{B}$ for $\boldsymbol{A} - \boldsymbol{B} \in \mathbb{S}_n^0$, with $\boldsymbol{A} \succ_L \boldsymbol{B}$ whenever $\boldsymbol{A} - \boldsymbol{B} \in \mathbb{S}_n^+$. Moreover, $\mathcal{C}(n)$ comprises the convex sets in $\mathbb{R}^n$ symmetric under reflection through $\mathbf{0} \in \mathbb{R}^n$. Operators $E(\boldsymbol{Y})$ and $V(\boldsymbol{Y})$ designate the expected vector and dispersion matrix for $\boldsymbol{Y} \in \mathbb{R}^n$; with var$(Y)$ as the variance on $\mathbb{R}^1$. We further require $\{\mu_r(Z) = E(Z^r); r = 1, 2\}$ as moments about $0 \in \mathbb{R}^1$, identifying $\kappa_2 = \mu_2(\hat{\beta}_1)$ in terms of a linear estimator $\hat{\beta}_1$ to be encountered subsequently. The comparative concentration of probability measures on $\mathbb{R}^n$ may be gauged on defining the measure $\mu(\cdot)$ to be *more peaked* about $\mathbf{0} \in \mathbb{R}^n$ than $\nu(\cdot)$, if and only if $\mu(A) \geq \nu(A)$ for every set $A \in \mathcal{C}(n)$, as in [2]. Specifically, the peakedness ordering for scale mixtures of Gaussian measures on $\mathbb{R}^n$ is tantamount to the stochastic ordering of their mixing distributions, as demonstrated in the Appendix.

2.2. *Special distributions*

Here *pdf* and *cdf* refer to probability density and cumulative distribution functions – for $\boldsymbol{Y} \in \mathbb{R}^n$, $\mathcal{L}(\boldsymbol{Y})$ designates its law of distribution and $G(\boldsymbol{y})$ its cdf – and *iid* refers to independent and identically distributed random elements. Distributions of note on $\mathbb{R}^1$ include the Gaussian law $N_1(\mu, \sigma^2)$, with parameters (μ, σ^2); non-central versions of student's $t(\nu, \lambda)$, and $t^2(\nu, \lambda)$, chi-squared $\chi^2(\nu, \lambda)$, and Snedecor–Fisher $F(\nu_1, \nu_2, \lambda)$ distributions, with $\{\nu, \nu_1, \nu_2\}$ as degrees of freedom and λ as a non-centrality parameter; and $G_0(\alpha, \beta)$ as the gamma distribution on $\mathbb{R}^1_+$ having parameters (α, β). In particular, $g_T(t; \nu, \lambda)$, $g_{T^2}(u; \nu, \lambda)$, $g_F(u; \nu_1, \nu_2, \lambda)$, and $g_0(u; \alpha, \beta)$ designate the densities corresponding to $t(\nu, \lambda)$, $t^2(\nu, \lambda)$, $F(\nu_1, \nu_2, \lambda)$, and $G_0(\alpha, \beta)$, respectively.

Futher, $N_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ designates the Gaussian law on $\mathbb{R}^n$ and $g_n(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$ its pdf, having location–scale parameters $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. Ensembles on $\mathbb{R}^n$, and mixtures over these, include the translation–scale mixtures

$$f_1(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma}, G_1) = \int_{-\infty}^{\infty} g_n(\boldsymbol{y}; \boldsymbol{\theta}(t), \boldsymbol{\Sigma}(t)) \mathrm{d}G_1(t) \tag{1}$$

and purely scale mixtures when $\boldsymbol{\theta} = \mathbf{0}$. Non-standard distributions for quadratic forms proceed conditionally on letting $\mathcal{L}(U|w)$ have the scaled gamma density $g_0(u; \alpha, w\beta) = (w\beta)^{-\alpha}u^{\alpha-1}\mathrm{e}^{-u/w\beta}/\Gamma(\alpha)$, then compounding these as

$$f_3(u; \alpha, \beta, G_3) = \frac{u^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \int_0^{\infty} w^{-\alpha}\mathrm{e}^{-x/w\beta}\mathrm{d}G_3(w) \tag{2}$$

with $G_3(w)$ as a cdf on $\mathbb{R}^1_+$.

Subsequent developments have links to exchangeable sequences. Consider $\{\mathbf{Z}_0, Z_1, Z_2, \ldots\}$ such that $\mathbf{Z}_0 \in \mathbb{R}^k$ is independent of $\{Z_1, Z_2, \ldots\}$, whereas $\{Z_1, Z_2, \ldots\}$ are iid on $\mathbb{R}^{\infty}$: Further, let $\psi : \mathbb{R}^{k+1} \to \mathbb{R}^1$, define $\{Y_i = \psi(\mathbf{Z}_0, Z_i); i = 1, 2, \ldots\}$, recall from DeFinetti's theorem that the sequence $\{Y_1, Y_2, \ldots\}$ is now *exchangeable* on $\mathbb{R}^{\infty}$, and infer that the joint distributions projected onto $\mathbb{R}^n$ are invariant under permutations. In short, $\{Y_1, \ldots, Y_n\}$ are distributed identically under second moments, having common values for the parameters (μ, σ^2, ρ).

3. Calibration

3.1. *Essentials*

Instruments are calibrated using outputs at successive inputs to determine a *calibration curve*, new readings are assigned values on the scale of measurements using the calibrated device, and periodic checks against a standard determine when recalibration is required. In this study, we utilize *direct assays* in which instrumental readings $\{X_i; 1 \le i \le n_0\}$ during calibration relate to observed measurements $\{U_i; 1 \le i \le n_0\}$ through $\{U_i = \beta_0 + \beta_1 X_i + \varepsilon_i; 1 \le i \le n_0\}$. For example, the octane rating (U) in the production of gasoline relates linearly to the percent of purity (X) in a specimen to be assayed. Octane numbers require expensive and time-consuming dynamic laboratory testing, whereas the percent purity is readily determined. Once calibrated, the octane number of a given specimen is determined vicariously from its percent purity. On the other hand, *indirect assays* proceed on reversing the roles of U_i and X_i during calibration. Models for calibration and their analyses have been debated by several authors; for a summary and early references, see [3]. Problems with moments and consistency remain to be resolved in indirect assays, but the technical issues between the two types differ mainly in detail. It is noteworthy that research has yet to address the principal issues undertaken here, namely, the irregularities in models and supporting analyses attributable to calibration.

3.2. *Error analysis*

Now, consider the calibrating model $\{U_i = \beta_0 + \beta_1 X_i + \varepsilon_i; 1 \le i \le n_0\}$ under Gauss–Markov assumptions, such that $\{\mathrm{var}(U_i) = \sigma_U^2; 1 \le i \le n_0\}$ and $(\hat{\beta}_0, \hat{\beta}_1)$ are least-squares estimators determining the empirical calibration line. Under *Gaussian calibration*, the calibration errors $\{\varepsilon_i; 1 \le i \le n_0\}$ comprise iid $N_1(0, \sigma_U^2)$ random variables. Subsequent readings $\{Z_1, \ldots, Z_n\}$, taken independently of $\{U_1, \ldots, U_{n_0}\}$, are then projected as the calibrated measurements $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \le i \le n\}$. In practice $\{Z_1, \ldots, Z_n\}$ often will have been discarded as redundant, or will have been converted directly without record, so that $\{Y_1, \ldots, Y_n\}$ remain to be analysed and interpreted. If we now suppose that $\boldsymbol{Z}' = [Z_1, \ldots, Z_n]$ have means $\boldsymbol{\mu}' = [\mu_1, \ldots, \mu_n]$ and second moments $V(\mathbf{Z}) = \Sigma = [\sigma_{ij}]$, independently of $(\hat{\beta}_0, \hat{\beta}_1)$, then the conditional moments of $\mathcal{L}(Y_1, \ldots, Y_n|\hat{\beta}_1)$ are found directly as follows.

LEMMA 1 *Suppose that* $\{Z_1, \ldots, Z_n\}$ *have means* $\{\mu_1, \ldots, \mu_n\}$ *and second moments* $V(\mathbf{Z}) = [\sigma_{ij}]$, *independently of* $(\hat{\beta}_0, \hat{\beta}_1)$,*and let* $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i;\ 1 \le i \le n\}$. *Then*

(i) $E(Y_i|\hat{\beta}_1) = E(\hat{\beta}_0|\hat{\beta}_1) + \hat{\beta}_1\mu_i$,
(ii) $\mathrm{var}(Y_i|\hat{\beta}_1) = \hat{\beta}_1^2\sigma_{ii} + \mathrm{var}(\hat{\beta}_0|\hat{\beta}_1)$, *and*
(iii) $\mathrm{cov}(Y_i, Y_j|\hat{\beta}_1) = \hat{\beta}_1^2\sigma_{ij} + \mathrm{var}(\hat{\beta}_0|\hat{\beta}_1)$.

If instead (β_0, β_1) were known, then $\{Y_i = \beta_0 + \beta_1 Z_i;\ 1 \le i \le n\}$ would be recovered without an error, in which case $E(Y_i) = \beta_0 + \beta_1\mu_i$ and $\mathrm{var}(Y_i) = \beta_1^2\mathrm{var}(Z_i)$. This ideal case serves as a reference against which a recovery subject to calibration errors may be gauged.

Expressions simplify when neither $E(\hat{\beta}_0|\hat{\beta}_1)$ nor $\mathrm{var}(\hat{\beta}_0|\hat{\beta}_1)$ depends on $\hat{\beta}_1$, so that $\mathrm{cov}(Y_i, Y_j|\hat{\beta}_1) = \hat{\beta}_1^2\sigma_{ij} + \mathrm{var}(\hat{\beta}_0)$, for example. This clearly holds under Gaussian calibration, provided that the calibrating readings $\{X_1, \ldots, X_{n_0}\}$ have been centred to $\{(X_1 - \overline{X}), \ldots, (X_{n_0} - \overline{X})\}$. This incurs no loss in generality, as subsequent readings $\{Z_1, \ldots, Z_n\}$ may be shifted by $\overline{X}$ units before projecting onto the scale of measurements. It then follows that $\hat{\beta}_0 = \overline{U}$; $\mathrm{var}(\hat{\beta}_0) = \sigma_0^2 = \sigma_U^2/n_0$; and $\mathrm{var}(\hat{\beta}_1) = \sigma_1^2 = \sigma_U^2/S_{xx}$, where $S_{xx} = \sum_{i=1}^{n_0}(X_i - \overline{X})^2$; so that $(\hat{\beta}_0, \hat{\beta}_1)$ are uncorrelated and thus independent under Gaussian calibration errors. We henceforth take the initial calibration to have been centred.

We next consider the conditional and unconditional properties of $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i;\ 1 \le i \le n\}$ for the general case that $E(\mathbf{Z}) = \boldsymbol{\mu}_Z \in \mathbb{R}^n$ and $V(\mathbf{Z}) = \boldsymbol{\Sigma} \in \mathbb{S}_n^+$, to be specialized subsequently. Clearly, the conditional means and dispersion parameters are $E(\boldsymbol{Y}|\hat{\beta}_1) = \beta_0\mathbf{1}_n + \hat{\beta}_1\boldsymbol{\mu}_Z = \boldsymbol{\mu}_Y(\hat{\beta}_1)$, say, and $V(\boldsymbol{Y}|\hat{\beta}_1) = \hat{\beta}_1^2\Sigma + \sigma_0^2\mathbf{1}_n\mathbf{1}_n' = \Xi(\hat{\beta}_1)$. Moreover, for the case that $\mathcal{L}(\mathbf{Z}) = N_n(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma})$ in addition to Gaussian calibration errors, the conditional distribution of $\boldsymbol{Y}$ is $\mathcal{L}(\boldsymbol{Y}|\hat{\beta}_1) = N_n(\boldsymbol{\mu}_Y(\hat{\beta}_1), \Xi(\hat{\beta}_1))$. Basic unconditional properties follow next.

THEOREM 1 *Consider the calibrated measurements* $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i;\ 1 \le i \le n\}$ *projected from readings* $\{Z_1, \ldots, Z_n\}$ *obtained independently of* $(\hat{\beta}_0, \hat{\beta}_1)$; *let* $\boldsymbol{Y} = \hat{\beta}_0\boldsymbol{1}_n + \hat{\beta}_1\mathbf{Z}$, *such that* $E(\mathbf{Z}) = \boldsymbol{\mu}_Z \in \mathbb{R}^n$ *and* $V(\mathbf{Z}) = \boldsymbol{\Sigma} \in \mathbb{S}_n^+$; *and let* $\sigma_0^2 = \mathrm{var}(\hat{\beta}_0)$ *and* $\sigma_1^2 = \mathrm{var}(\hat{\beta}_1)$. *Then the unconditional moments* $E(\boldsymbol{Y}) = \boldsymbol{\mu}_Y$ *and* $V(\boldsymbol{Y}) = \Xi$ *of* $\mathcal{L}(\boldsymbol{Y})$ *are given by*

(i) $\boldsymbol{\mu}_Y = \beta_0\boldsymbol{1}_n + \beta_1\boldsymbol{\mu}_Z$, *and*
(ii) $\boldsymbol{\Xi} = \kappa_2\boldsymbol{\Sigma} + \sigma_0^2\boldsymbol{1}_n\boldsymbol{1}_n' + \sigma_1^2\boldsymbol{\mu}_Z\boldsymbol{\mu}_Z'$, *with* $\kappa_2 = \mu_2(\hat{\beta}_1) = \sigma_1^2 + \beta_1^2$.
(iii) *Moreover, if* $\mathcal{L}(\mathbf{Z}) = N_n(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma})$ *in addition to Gaussian calibration errors, then the unconditional joint density of the elements of* $\boldsymbol{Y}$ *is the translation–scale mixture*

$$f_1(\mathbf{y}; \boldsymbol{\mu}_Y, \boldsymbol{\Xi}, G_1) = \int_{-\infty}^{\infty} g_n(\mathbf{y}; \boldsymbol{\mu}(t), \boldsymbol{\Xi}(t))\mathrm{d}G_1(t) \tag{3}$$

as in Equation (1), *with* $\boldsymbol{\mu}(t) = \beta_0\boldsymbol{1}_n + t\boldsymbol{\mu}_Z$, $\boldsymbol{\Xi}(t) = t^2\boldsymbol{\Sigma} + \sigma_0^2\boldsymbol{1}_n\boldsymbol{1}_n'$, *and with mixing distribution* $G_1(t) = N_1(\beta_1, \sigma_1^2)$.

Proof Conclusion (i) follows directly through deconditioning. Conclusion (ii) follows using $\mathrm{var}(Y_i) = E_{\hat{\beta}_1}[\mathrm{var}(Y_i \mid \hat{\beta}_1)] + \mathrm{var}_{\hat{\beta}_1}(E(Y_i \mid \hat{\beta}_1)]$ for variances and

$$\mathrm{cov}(Y_i, Y_j) = E_{\hat{\beta}_1}[\mathrm{cov}(Y_i, Y_j \mid \hat{\beta}_1)] + \mathrm{cov}_{\hat{\beta}_1}[E(Y_i \mid \hat{\beta}_1), E(Y_j \mid \hat{\beta}_1)]$$

for covariances. Conclusion (iii) follows since $\boldsymbol{Y}$ is a linear function of $(\mathbf{Z}, \hat{\beta}_0)$ with $\hat{\beta}_1$ fixed, so that $\mathcal{L}(\boldsymbol{Y} \mid \hat{\beta}_1) = N_n(\boldsymbol{\mu}_y(\hat{\beta}_1), \boldsymbol{\Xi}(\hat{\beta}_1))$, as noted, and then mixing over the distribution of the conditioning variable. ■

The above-mentioned results are basic. We next specialize them as appropriate for specific experimental settings encountered routinely in practice.

4. Topics in inference

Induced dependencies and other model irregularities violate the tenets of conventional data analysis as noted, specifically, in estimation and hypothesis testing. We focus on normal-theory procedures, as the independence typically required by non-parametric competitors is conspicuously absent. The following sections specialize earlier findings, as they apply in a single sample, and in one-way comparative experiments.

4.1. *Single sample*

Consider $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \le i \le n\}$ such that the elements of $\boldsymbol{Z} = [Z_1, \ldots, Z_n]'$ satisfy $\{E(Z_i) = \mu_Z; 1 \le i \le n\}$ and $V(\boldsymbol{Z}) = \sigma_Z^2 \boldsymbol{I}_n$. If in addition $\{Z_1, \ldots, Z_n\}$ are iid, then $\mathcal{L}(Y_1, \ldots, Y_n)$ is exchangeable, as noted earlier. We are concerned not only with properties of the joint distribution $\mathcal{L}(\boldsymbol{Y})$, but also of $(\overline{Y}, S_Y^2, t_0)$ as the sample mean, the sample variance, and Student's statistic $t_0 = n^{1/2}(\overline{Y} - \mu_Y^0)/S_Y$, as well as the ordinary residuals $\{R_i = (Y_i - \overline{Y}); 1 \le i \le n\}$. From Lemma 1 and Theorem 1, we now have the conditional and unconditional values $\{E(Y_i \mid \hat{\beta}_1) = \beta_0 + \hat{\beta}_1 \mu_Z; 1 \le i < n\}$; $\{E(Y_i) = \mu_Y = \beta_0 + \beta_1 \mu_Z; 1 \le i \le n\}$; and $\mathrm{V}(\boldsymbol{Y} \mid \hat{\beta}_1) = \boldsymbol{\Xi}(\hat{\beta}_1) = (\hat{\beta}_1^2 \sigma_Z^2 + \sigma_0^2)[(1-\rho)\boldsymbol{I}_n + \rho \mathbf{1}_n \mathbf{1}_n']$, with $\rho = \sigma_0^2/(\hat{\beta}_1^2 \sigma_Z^2 + \sigma_0^2)$. The unconditional variances are homoscedastic, namely $\{\mathrm{var}(Y_i) = \sigma_Y^2 = \kappa_2 \sigma_Z^2 + \sigma_0^2 + \sigma_1^2 \mu_Z^2; 1 \le i \le n\}$. Essential findings follow, where it is seen that S_Y^2 may grossly underestimate the actual measurement variance σ_Y^2, and that structural difficulties becloud both the small-sample and the asymptotic properties of $\overline{Y}_n = (Y_1 + Y_2 + \cdots + Y_n)/n$.

THEOREM 2 *Let* $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \le i \le n\}$ *be calibrated measurements from* $\{Z_1, \ldots, Z_n\}$, *such that* $E(\boldsymbol{Z}) = \mu_Z \boldsymbol{I}_n$ *and* $V(\boldsymbol{Z}) = \sigma_Z^2 \boldsymbol{I}_n$ *independently of* $(\hat{\beta}_0, \hat{\beta}_1)$, *and consider the sample quantities* $(\overline{Y}_n, S_Y^2)$, *together with the ordinary residuals* $\{R_i = (Y_i - \overline{Y}); 1 \le i \le n\}$. *Then*

(i) $\overline{Y}_n$ *is unbiased but inconsistent for estimating* $\mu_Y = \beta_0 + \beta_1 \mu_Z$,
(ii) $E(S_Y^2) = \kappa_2 \sigma_Z^2 = \sigma_Z^2(\sigma_1^2 + \beta_1^2)$, *and*
(iii) $\{E(R_i) = 0; 1 \le i \le n\}$.

Proof The unbiasedness of $\overline{Y}_n$ follows routinely, and its variance from $\mathrm{var}(n^{-1}\mathbf{1}_n'\boldsymbol{Y}) = \Xi_n$ with

$$\begin{aligned} \Xi_n &= n^{-2}\mathbf{1}_n'[\kappa_2 \sigma_Z^2 \mathbf{I}_n + (\sigma_0^2 + \sigma_1^2 \mu_Z^2)\mathbf{1}_n \mathbf{1}_n']\mathbf{1}_n \\ &= n^{-1}\kappa_2 \sigma_Z^2 + (\sigma_0^2 + \sigma_1^2 \mu_Z^2). \end{aligned} \tag{4}$$

Since $\lim_{n\to\infty} \mathrm{var}(\overline{Y}_n) = (\sigma_0^2 + \sigma_1^2 \mu_Z^2) > 0$, it follows that its limit distribution is non-degenerate at μ_Y, so that $\overline{Y}_n$ is consistent neither in probability, nor in mean square, nor almost surely, as asserted. Conclusion (ii) follows from evaluating the expected value of the quadratic form $(n-1)S_Y^2 = \boldsymbol{Y}'\boldsymbol{B}_n\boldsymbol{Y}$ as $E[(n-1)S_Y^2] = \mathrm{tr}(\boldsymbol{B}_n V(\boldsymbol{Y})) + \boldsymbol{\mu}_Y'\boldsymbol{B}_n\boldsymbol{\mu}_Y$. Details are

$$\begin{aligned} E[(n-1)S_Y^2] &= \mathrm{tr}\left(\boldsymbol{B}_n\left[k_2\sigma_Z^2\boldsymbol{I}_n + \left(\sigma_0^2 + \sigma_1^2\mu_Z^2\right)\mathbf{1}_n\mathbf{1}_n'\right]\boldsymbol{B}_n\right) + \boldsymbol{\mu}_Y'\boldsymbol{B}_n\boldsymbol{\mu}_Y \\ &= (n-1)k_2\sigma_Z^2 \end{aligned}$$

where $\boldsymbol{\mu}_Y'\boldsymbol{B}_n\boldsymbol{\mu}_Y = (\beta_0 + \beta_1\mu_Z)^2\mathbf{1}_n'\boldsymbol{B}_n\mathbf{1}_n = 0$, since $\boldsymbol{B}_n$ is idempotent and $\boldsymbol{B}_n\mathbf{1}_n = \mathbf{0}$. Conclusion (iii) follows directly, to complete our proof. ■

The following consequences may be noted.

- Conclusion (i) appears to dash the usual expectation that lengths of $(1-\alpha)$ confidence intervals for μ_Y will decrease at the rate $O(n^{-1/2})$.
- The sample variance S_Y^2 underestimates the actual variance σ_Y^2. The bias is $E(S_Y^2)-\sigma_Y^2 = -(\sigma_0^2+\sigma_1^2\mu_Z^2)$.
- This bias increases with decreasing precision in estimating the calibration line, and with increasing $|\mu_Z|$ and thus $|\mu_Y|$.
- On the other hand, the expectation $E(S_Y^2)=\kappa_2\sigma_Z^2$, with $\kappa_2=\mu_2(\hat{\beta}_1)$, may be compared with the ideal variance, $\mathrm{var}(Y_i)=\beta_1^2\sigma_Z^2$, attained under linear calibration with known parameters (β_0,β_1).

We have seen how unconditional moments of calibrated measurements depend on those of the conditioning variable $\hat{\beta}_1$. It remains to examine effects of the fitted calibration line on unconditional distributions, including those of various sample statistics. Recall from Theorem 2 and its proof that $E(\overline{Y})=\mu_Y=\beta_0+\beta_1\mu_Z$ and $\mathrm{var}(\overline{Y})=n^{-1}k_2\sigma_Z^2+(\sigma_0^2+\sigma_1^2\mu_Z^2)$. To invoke expression (1) and its special case at $\boldsymbol{\theta}=\mathbf{0}$, under Gaussian assumptions, we have $G_1(\hat{\beta}_1)=N_1(\beta_1,\sigma_1^2)$, together with $G_2(\hat{\beta}_1^2;\lambda)$, such that $\mathcal{L}(\hat{\beta}_1^2/\sigma_1^2)=\chi^2(1,\lambda)$ with $\lambda=\beta_1^2/\sigma_1^2$. Basic unconditional distributions follow next as mixtures.

THEOREM 3 *Let* $\{Y_i=\hat{\beta}_0+\hat{\beta}_1Z_i;\,1\le i\le n\}$ *be calibrated measurements*; *suppose that* $\{Z_1,\ldots,Z_n\}$ *are iid* $N_1(\mu_Z,\sigma_Z^2)$ *idependently of* $(\hat{\beta}_0,\hat{\beta}_1)$ *under Gaussian calibration*; *and consider the sample quantities* $(\overline{Y}_n, S_Y^2, t_0^2)$, *together with the ordinary residuals* $\{R_i=(Y_i-\overline{Y});\,1\le i\le n\}$, *where* $t_0^2=n(\overline{Y}-\mu_Y^0)^2/S_Y^2$ *for testing* $H_0:\mu_Y=\mu_Y^0$ *against* $H_1:\mu_Y\ne\mu_Y^0$. *Then the unconditional properties are a follows.*

(i) *The unconditional density of* $\mathcal{L}(\overline{Y})$ *is the translation–scale mixture*

$$f_1(u;\mu_Y,\Xi_n,G_1)=\int_{-\infty}^{\infty} g_1(u;\mu(t),\Xi_n(t))\mathrm{d}G_1(t) \tag{5}$$

with mixing distribution $G_1(\hat{\beta}_1)=N_1(\beta_1,\sigma_1^2)$ *as in Equation* (1) *for the case* $n=1$, *where* $\mu(t)=\beta_0+t\mu_Z$ *and* $\Xi_n(t)=(t^2\sigma_Z^2/n+\sigma_0^2)$, *and* Ξ_n *is defined in Equation* (4).

(ii) *The joint density of residuals* $R=[R_1,\ldots,R_n]'$ *is given by*

$$f_n(\boldsymbol{r};\mathbf{0},\sigma_Z^2\boldsymbol{B}_n,G_2)=\int_0^{\infty} g_n(\boldsymbol{r};\mathbf{0},t\sigma_Z^2\boldsymbol{B}_n)\mathrm{d}G_2(t) \tag{6}$$

with mixing distribution $G_2(\hat{\beta}_1^2;\lambda)=\chi^2(1,\lambda)$ *and* $\lambda=\beta_1^2/\sigma_1^2$.

(iii) *The joint distribution* $\mathcal{L}(R_1,\ldots,R_n)$ *increases in peakedness about* $0\in\mathbb{R}^n$ *with decreasing* λ.

(iv) $\mathcal{L}(\upsilon S_Y^2/\sigma_Z^2)$ *has the density* $f_0(u;\upsilon/2,2,G_2)$ *as in Equation* (2), *with* $\upsilon=n-1$, *mixing distribution* $G_2(\hat{\beta}_1^2;\lambda)=\chi^2(1,\lambda)$ *and* $\lambda=\beta_1^2/\sigma_1^2$, *and* $E(S_Y^2)=\sigma_Z^2(\sigma_1^2+\beta_1^2)$.

(v) *The distribution* $\mathcal{L}(\upsilon S_Y^2/\sigma_Z^2)$ *increases stochastically with* $\lambda=\beta_1^2/\sigma_1^2$.

(vi) *The unconditional density of* $t_0^2=n(\overline{Y}-\mu_Y^0)^2/S_Y^2$ *is given as the mixture*

$$g(u;\upsilon,\delta,G_2)=\int_0^{\infty} g_{T^2}\left(u;\upsilon,\frac{\delta}{t}\right)\mathrm{d}G_2(t)$$

with mixing distribution $G_2(\hat{\beta}_1^2;\lambda)$, *where* $\upsilon=n-1$, $\delta=(\mu_Y-\mu_Y^0)^2/\sigma_Z^2$, *and* $\lambda=\beta_1^2/\sigma_1^2$.

(vii) *The unconditional cdf of* $\mathcal{L}(t_0^2)$ *increases stochastically with increasing* $\delta=(\mu_Y-\mu_Y^0)^2/\sigma_Z^2$ *for fixed* $\lambda=\beta^2/\sigma_1^2$, *and for fixed* δ, *it decreases stochastically with increasing* λ.

Proof The conditional distribution of note is $\mathcal{L}(\overline{Y}|\hat{\beta}_1) = N_1(\mu(\hat{\beta}_1), \Xi_n(\hat{\beta}_1))$, where $\mu(\hat{\beta}_1) = \beta_0 + \hat{\beta}_1\mu_Z$ and $\Xi_n(\hat{\beta}_1) = (\hat{\beta}_1^2\sigma_Z^2/n + \sigma_0^2)$. Its unconditional density thus is $f_1(u; \mu_Y, \Xi_n, G_1)$ from Equation (1), to give conclusion (i) with mixing distribution as asserted. Now, observe that $\boldsymbol{R} = \boldsymbol{B}_n\boldsymbol{Y}$ and $\boldsymbol{B}_n\Xi(\hat{\beta}_1)\boldsymbol{B}_n = \boldsymbol{B}_n(\hat{\beta}_1^2\sigma_Z^2\boldsymbol{I}_n + \sigma_0^2\mathbf{1}_n\mathbf{1}_n')\boldsymbol{B}_n = \hat{\beta}_1^2\sigma_Z^2\boldsymbol{B}_n$ since $\boldsymbol{B}_n$ is idempotent and $\boldsymbol{B}_n\mathbf{1}_n = \mathbf{0}$. We infer conditionally that $\mathcal{L}(R|\hat{\beta}_1) = N_n(\mathbf{0}, \hat{\beta}_1^2\sigma_Z^2\mathbf{B}_n)$, since $\boldsymbol{B}_n\boldsymbol{\mu}(\hat{\beta}_1) = (\beta_0 + \hat{\beta}_1\mu_Z)\boldsymbol{B}_n\mathbf{1}_n = \mathbf{0}$. The unconditional distribution is the scale mixture as in conclusion (ii), with $G_2(\hat{\beta}_1^2, \lambda) = \chi^2(1, \lambda)$ as the mixing distribution over $[0, \infty]$. Conclusion (iii) follows from Lemma A1 of the Appendix, as the mixing distribution $\mathcal{L}(\hat{\beta}_1^2/\sigma_1^2) = \chi^2(1, \lambda)$ increases stochastically with $\lambda = \beta_1^2/\sigma_1^2$. Now, observe that $(n-1)S_Y^2 = \boldsymbol{R}'\boldsymbol{R}$, so that $\mathcal{L}(\boldsymbol{R}'\boldsymbol{R}/\hat{\beta}_1\sigma_Z^2|\hat{\beta}_1^2) = \chi^2(n-1, 0)$. It follows that $\mathcal{L}[(n-1)S_Y^2/\sigma_Z^2|\hat{\beta}_1^2]$ is a central chi-squared variate scaled by $\hat{\beta}_1^2$. On identifying $(n-1)S_Y^2/\sigma_Z^2$ with U and $\hat{\beta}_1^2$ with w in developments leading to Equation (2), we thus establish conclusion (iv) on specializing from gamma to chi-squared distributions. Conclusion (v) follows directly from (iii) since the set $A_t = \{\boldsymbol{R} \in \mathbb{R}^n : \boldsymbol{R}'\boldsymbol{R} \leq t\}$ is convex and symmetric in $\mathbb{R}^n$, whereas $P(A_t) = P[(n-1)S_Y^2 \leq t\sigma_Z^2]$.

To see conclusion (vi), recall the affine invariance of $t^2 = n(\overline{Y} - \mu_Y^0)^2/S_Y^2$ for testing $H_0 : \mu_Y = \mu_Y^0$ under $\{Y_i \to a + bZ_i; 1 \leq i \leq n\}$, namely, $t^2 = n[(\overline{Y} - \mu_Y) + (\mu_Y - \mu_Y^0)]^2/S_Y^2 = n\{[(a + b\overline{Z}) - (a + b\mu_Z)] + b(\mu_Y - \mu_Y^0)/b\}^2/b^2S_Z^2 = n[(\overline{Z} - \mu_Z) + (\mu_Y - \mu_Y^0)/b]^2/S_Z^2$. Clearly $\mathcal{L}(t^2) = t^2(v, \delta)$, with $v = n-1$ and $\delta = (\mu_Y - \mu_Y^0)^2/b^2\sigma_Z^2$, independently of a. We next apply these facts conditionally, given $(\hat{\beta}_0, \hat{\beta}_1)$, on replacing (a, b) with $(\hat{\beta}_0, \hat{\beta}_1)$, to infer that $\mathcal{L}(t_0^2|\hat{\beta}_0, \hat{\beta}_1) = \mathcal{L}(t_0^2|\hat{\beta}_1) = t^2(v, \delta(\hat{\beta}_1))$, where $\delta(\hat{\beta}_1) = (\mu_Y - \mu_Y^0)^2/\hat{\beta}_1^2\sigma_Z^2$, independently of $\hat{\beta}_0$. It follows that the unconditional distribution of t_0^2 has the mixture density $g(u; v, \delta, G_2) = \int_0^\infty g_{T^2}(u; v, \delta/u)\mathrm{d}G_2(u)$ with mixing distribution $G_2(\hat{\beta}_1^2; \lambda)$ as asserted, where $\delta = (\mu_Y - \mu_Y^0)^2/\sigma_Z^2$ and $\lambda = \hat{\beta}_1^2/\sigma_1^2$. Therefore, the test for $H_0 : \mu_Y = \mu_Y^0$ against $H_1 : \mu_Y \neq \mu_Y^0$ rejects at level α for $t_0^2 > c_\alpha^2$; moreover, the conditional cdf $\mathcal{L}(t_0^2|\hat{\beta}_1)$ increases stochastically with $\delta(\hat{\beta}_1)$, pointwise for each fixed $\hat{\beta}_1$ from standard properties of non-central t^2 distributions. It follows that the unconditional cdf increases stochastically with increasing δ under mixing. That the unconditional cdf decreases stochastically with increasing λ with δ held fixed, follows unconditionally through mixing as in the proof for conclusion (v), to complete our proof. ■

Note from conclusion (iv) that $E(S_Y^2) = \sigma_Z^2(\sigma_1^2 + \beta_1^2)$. This may be compared with the ideal case $\mathrm{var}(Y_i) = \beta_1^2\sigma_Z^2$, where (β_0, β_1) are known. It is instructive to re-examine the unconditional properties of this section if we retain the homogeneity of variances of $\{Z_1, \ldots, Z_n\}$, but assume instead that means are not, i.e., that $E(Z) = \mu = [\mu_1, \ldots, \mu_n]'$. Therefore, we see that $\{\mathrm{var}(Y_i) = \kappa_2\sigma_Z^2 + \sigma_0^2 + \sigma_1^2\mu_i^2; 1 \leq i \leq n\}$. We have the curious finding that calibrated measurements in a single sample will have heterogeneous variances under heterogeneous means, despite the homogeneity of variances of $\{Z_1, \ldots, Z_n\}$.

4.2. *One-way experiments*

Clearly $\{Y_1, \ldots, Y_n\}$ and $\{Z_1, \ldots, Z_n\}$ have the same experimental structure, here a one-way experiment comprising k samples of sizes $\{n_1, \ldots, n_k\}$, with $n_1 + \cdots + n_k = n$ In keeping with conventional notation, partition $\boldsymbol{Z}' = [Z_1, \ldots, Z_n]$ as $[\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_n']$ such that $\{\boldsymbol{Z}_i' = [Z_{i1}, \ldots, Z_{in_i}]; 1 \leq i \leq k\}$; similarly for $Y' = [\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_k']$, with $\{\boldsymbol{Y}_i' = [Y_{i1}, \ldots, Y_{in_i}]; 1 \leq i \leq k\}$; and suppose that $\{E(Z_{ij}) = \mu_i; 1 \leq j \leq n_i\}$ and $\{\mathrm{var}(Z_{ij}) = \omega_i^2; 1 \leq j \leq n_i\}$. Accordingly, take $\mu_Z = [\mu_1 1_{n_1}', \ldots, \mu_k 1_{n_k}']'$ and $V(\boldsymbol{Z}) = \mathrm{diag}(\omega_1^2\boldsymbol{I}_{n_1}, \ldots, \omega_k^2\boldsymbol{I}_{n_k}) = \boldsymbol{D}(\omega_1^2, \ldots, \omega_k^2)$, say. Specializing from Section 3, we have the conditional moments $E(\boldsymbol{Y}|\hat{\beta}_1) = \boldsymbol{\mu}_Y(\hat{\beta}_1) = \beta_0\mathbf{1}_n + \hat{\beta}_1[\mu_1\mathbf{1}_{n_1}', \ldots, \mu_k\mathbf{1}_{n_k}']'$ and $V(\boldsymbol{Y}|\hat{\beta}_1) = \Xi(\hat{\beta}_1) = \hat{\beta}_1^2\mathrm{diag}(\omega_1^2\boldsymbol{I}_{n_1}, \ldots, \omega_k^2\boldsymbol{I}_{n_k}) + \sigma_0^2\mathbf{1}_n\mathbf{1}_n'$, together with $\mathcal{L}(\boldsymbol{Y}|\hat{\beta}_1) = N_n(\boldsymbol{\mu}_Y(\hat{\beta}_1), \Xi(\hat{\beta}_1))$ under Gaussian errors. Moreover, unconditional moments

are $E(\boldsymbol{Y}) = \boldsymbol{\mu}_Y = \beta_0 \mathbf{1}_n + \beta_1[\mu_1 \mathbf{1}'_{n_1}, \ldots, \mu_k \mathbf{1}'_{n_k}]'$ and $V(\boldsymbol{Y}) = \Xi = k_2 \mathrm{diag}(\omega_1^2 \boldsymbol{I}_{n_1}, \ldots, \omega_k^2 \boldsymbol{I}_{n_k}) + \sigma_0^2 \mathbf{1}_n \mathbf{1}'_n + \sigma_1^2 \boldsymbol{M}$, where $\boldsymbol{M} = [\boldsymbol{M}_{ij}] = \boldsymbol{\mu}_Z \boldsymbol{\mu}'_Z$ has the block structure $\boldsymbol{M}_{ij} = \mu_i \mu_j \mathbf{1}_{n_i} \mathbf{1}'_{n_j}$. In particular, for typical calibrated measurements in sample i of the k samples, the conditional and unconditional means are $\{E(Y_{ij}|\hat{\beta}_1) = \beta_0 + \hat{\beta}_1 \mu_i;\ 1 \le j \le n_i\}$ and $\{E(Y_{ij}) = \beta_0 + \beta_1 \mu_i;\ 1 \le j \le n_i\}$, whereas the corresponding variances are $\{\mathrm{var}(Y_{ij}|\hat{\beta}_1) = \hat{\beta}_1^2 \omega_i^2 + \sigma_0^2;\ 1 \le j \le n_i\}$ and $\{\mathrm{var}(Y_{ij}) = \kappa_2 \omega_i^2 + \sigma_0^2 + \sigma_1^2 \mu_i^2;\ 1 \le j \le n_i\}$. We are concerned with the dual issues of the homogeneity of means, and of the homogeneity of variances, across the k samples. Clearly the induced irregularities are artifacts of the calibration process, rather than consequences of the experimental structure itself. It remains to determine unintended effects of calibration on conventional comparisons among the means and the variances.

In contrast to conventional one-way experiments, where homoscedasticity can be checked regardless of heterogeneity among the k population means, under calibration we see that homogeneity of the unconditional variances is possible only in unusual circumstances. Specifically, homoscedasticity holds unconditionally if and only if, for every pair (ω_i^2, μ_i^2) and (ω_j^2, μ_j^2), that $(\omega_i^2 - \omega_j^2) = c(\mu_j^2 - \mu_i^2)$ with $c = \sigma_1^2/\kappa_2$.

Now, consider the transformations $T_1(\boldsymbol{Y})$, $T_2(\boldsymbol{Y})$, and $T_3(\boldsymbol{Y})$ such that $T_1(\boldsymbol{Y}) = \overline{\boldsymbol{Y}} = [\overline{Y}, \ldots, \overline{Y}_k]'$ comprise the k sample means; $T_2(\boldsymbol{Y}) = \boldsymbol{R}' = [\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_k]'$ consists of the ordinary within-sample residuals, with $\boldsymbol{R}_i = \boldsymbol{B}_{n_i} \boldsymbol{Y}_i$ and $\boldsymbol{B}_{n_i} = (\boldsymbol{I}_{n_i} - n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i})$; and $T_3(\boldsymbol{Y}) = [\nu_1 S_1^2, \ldots, \nu_k S_k^2]'$ are the residual sums of squares, i.e., $\{\nu_i S_i^2 = \boldsymbol{R}'_i \boldsymbol{R}_i = \boldsymbol{Y}'_i \boldsymbol{B}_{n_i} \boldsymbol{Y}_i;\ 1 \le i \le k\}$, with $\nu_i = n_i - 1$. Basic properties may be summarized as follows.

THEOREM 4 *Consider the calibrated measurements* $\boldsymbol{Y}' = [\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_k]$ *corresponding to* $\boldsymbol{Z}' = [\boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_k]$ *such that* $E(\boldsymbol{Z}) = [\mu_1 \boldsymbol{1}'_{n_1}, \ldots, \mu_k \boldsymbol{1}'_{n_k}]'$ *and* $V(\boldsymbol{Z}) = \mathrm{Diag}(\omega_1^2 \boldsymbol{I}_{n_1}, \ldots, \omega_k^2 \boldsymbol{I}_{n_k})$, *and let* $T_1(\boldsymbol{Y}) = \overline{\boldsymbol{Y}} = [\overline{Y}_1, \ldots, \overline{Y}_k]'$, $T_2(\boldsymbol{Y}) = [\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_k]$, *and* $T_3(\boldsymbol{Y}) = [\nu_1 S_1^2, \ldots, \nu_k S_k^2]'$, *with* $\{\nu_i = n_i - 1;\ 1 \le i \le k\}$. *Moreover, a Gaussian model asserts that* $\{(Z_{ij} - \mu_i)/\omega_i;\ 1 \le j \le n_i,\ 1 \le i \le k\}$ *are iid* $N_1(0, 1)$ *random variates independently of* $(\hat{\beta}_0, \hat{\beta}_1)$ *under Gaussian calibration.*

(i) *The conditional and unconditional moments of* $T_1(\boldsymbol{Y}) = \overline{\boldsymbol{Y}}$ *are given by* $E(\overline{\boldsymbol{Y}}|\hat{\beta}_1) = \boldsymbol{\theta}(\hat{\beta}_1) = \beta_0 \boldsymbol{I}_k + \hat{\beta}_1 \boldsymbol{\mu}$, $E(\overline{\boldsymbol{Y}}) = \boldsymbol{\theta} = \beta_0 \mathbf{1}_k + \beta_1 \boldsymbol{\mu}$, $V(\overline{\boldsymbol{Y}}|\hat{\beta}_1) = \Xi_1(\hat{\beta}_1) = \hat{\beta}_1^2 \,\mathrm{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 \boldsymbol{I}_k \boldsymbol{I}'_k$, *and* $V(\overline{\boldsymbol{Y}}) = \Xi_1 = \kappa_2 \,\mathrm{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 \mathbf{1}_k \mathbf{1}'_k + \sigma_1^2 \boldsymbol{\mu}\boldsymbol{\mu}'$, *where* $\boldsymbol{\mu}' = [\mu_1, \ldots, \mu_k]$.

(ii) *Under Gaussian models, the unconditional density of* $\mathcal{L}(\overline{\boldsymbol{Y}})$ *is the translation–scale mixture*

$$f_k(u; \boldsymbol{\theta}, \Xi_1, G_1) = \int_{-\infty}^{\infty} g_k(u; \boldsymbol{\theta}(t), \Xi_1(t)) \mathrm{d}G_1(t) \tag{7}$$

with mixing distribution $G_1(\hat{\beta}_1) = N_1(\beta_1, \sigma_1^2)$ *as in Equation* (1), *where* $\boldsymbol{\theta}(t) = \beta_0 \mathbf{1}_k + t\boldsymbol{\mu}$ *and* $\Xi_1(t) = t^2 \,\mathrm{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 \mathbf{1}_k \mathbf{1}'_k$.

(iii) *The conditional and unconditional moments of the residuals are* $E(\boldsymbol{R} \mid \hat{\beta}_1) = E(\boldsymbol{R}) = \boldsymbol{0}$, $V(\boldsymbol{R}|\hat{\beta}_1) = \boldsymbol{\Xi}_2(\hat{\beta}_1) = \hat{\beta}_1^2 \,\mathrm{diag}(\omega_1^2 \boldsymbol{B}_{n_1}, \ldots, \omega_k^2 \boldsymbol{B}_{n_k})$, *and*

$$V(\boldsymbol{R}) = \boldsymbol{\Xi}_2 = \kappa_2 \,\mathrm{diag}(\omega_1^2 \boldsymbol{B}_{n_1}, \ldots, \omega_k^2 \boldsymbol{B}_{n_k}).$$

(iv) *Under Gaussian models, the joint density of residuals* $\boldsymbol{R} = [\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_k]'$ *is given by* $f_2(\boldsymbol{r}; \boldsymbol{0}, \boldsymbol{\Xi}_2, G_2)$ *as in Equation* (6), *with mixing distribution* $G_2(\hat{\beta}_1^2; \lambda)$ *and* $\lambda = \beta_1^2/\sigma_1^2$.

(v) *Under Gaussian models, the joint density of elements of* $[\nu_1 S_1^2/\omega_1^2, \ldots, \nu_k S_k^2/\omega_k^2]'$ *is given by*

$$f(\boldsymbol{u}; \nu_1, \ldots, \nu_k) = \int_0^{\infty} \prod_{i=1}^{k} g_0\left(u_i; \frac{\nu_i}{2}, 2w\right) \mathrm{d}G_2(w)$$

with $\nu_i = n_i - 1$ *and* $g_0(u; \alpha, w\beta) = (w\beta)^{-\alpha} u^{\alpha-1} \mathrm{e}^{-u/w\beta} / \Gamma(\alpha)$, *having the mixing distribution* $G_2(\hat{\beta}_1^2; \lambda)$ *with* $\lambda = \beta_1^2/\sigma_1^2$.

Proof Arguments follow step by step as in the proofs given in Section 4.1. The details differ, but proceed similarly on noting that $\overline{\boldsymbol{Y}} = \text{diag}(n_1^{-1}\mathbf{1}_{n_1}, \ldots, n_k^{-1}\mathbf{1}'_{n_k})\boldsymbol{Y} = \boldsymbol{L}_1'\boldsymbol{Y}$, say, whereas $\boldsymbol{R} = \text{diag}(\boldsymbol{B}_{n_1}, \ldots, B_{n_k})\boldsymbol{Y} = \boldsymbol{L}_2'\boldsymbol{Y}$. The conditional and unconditional moments follow directly as linear functions, together with the idempotencies of $\{\boldsymbol{B}_{n_i}; 1 \le i \le k\}$ and the annihilations achieved through $\{\boldsymbol{B}_{n_i}\mathbf{1}_{n_i} = \mathbf{0}; 1 \le i \le k\}$. A Gaussian model for $\mathcal{L}(\boldsymbol{Z})$ and Gaussian errors during calibration give conditional Gaussian laws for $\mathcal{L}(\overline{\boldsymbol{Y}} \mid \hat{\beta}_1)$ and $\mathcal{L}(\boldsymbol{R} \mid \hat{\beta}_1)$, whereas the unconditional distributions are mixtures as in Section 2.2. Moreover, $\{S_1^2, \ldots, S_k^2\}$, are conditionally independent given $\hat{\beta}_1$. As in the proof for Theorem 3(iv), the marginal density of $\mathcal{L}(\nu_i S_i^2/\omega_i^2 \mid \hat{\beta}_1)$ is the scaled chi-squared density $g_0(u_i; \nu_i/2, 2\omega)$ as defined preceding Equation (2), with $w = \hat{\beta}_1$. Their unconditional density now follows on mixing as in Section 2.2, as asserted in conclusion (v), to complete our proof. ■

It is essential to examine effects of calibration in comparing variances across the k groups, typically based on $\{S_1^2, \ldots, S_k^2\}$. In the ideal case where (β_0, β_1) are known, we would have $\{\text{var}(Y_{ij}) = \beta_1^2\omega_i^2; 1 \le i \le k\}$, so that homoscedasticity across groups for measurements $\{Y_{ij}\}$ would be tantamount to that for $\{Z_{ij}\}$. Under calibration errors, however, Theorem 2(ii) shows that S_i^2 underestimates $\text{var}(Y_{ij}) = \kappa_2\omega_i^2 + \sigma_0^2 + \sigma_1^2\mu_i^2$, the amount of bias, $B_i = -(\sigma_0^2 + \sigma_1^2\omega_i^2)$, being an artifact of the calibration process itself. Accordingly, it is germane to examine homogeneity among the expected values $\{\kappa_2\omega_1^2, \ldots, \kappa_2\omega_k^2\}$ of $\{S_1^2, \ldots, S_k^2\}$. To these ends, let $T_4(S_1^2, \ldots, S_k^2)$ be any scale-invariant statistic based on the sample variances from the measurements $\{Y_{ij}; 1 \le j \le n_i, 1 \le i \le k\}$. The following result is basic.

THEOREM 5 *Let $\{S_1^2, \ldots, S_k^2\}$ be within-sample variances from the calibrated measurements $\{Y_{ij}; 1 \le j \le n_i, 1 \le i \le k\}$ in a one-way experiment; let $T_4(S_1^2, \ldots, S_k^2)$ be any scale-invariant statistic; and consider a Gaussian model where $\{(Z_{ij} - \mu_i)/\omega_i; 1 \le j \le n_i, 1 \le i \le k\}$ are iid $N_1(0, 1)$ random variables independently of $(\hat{\beta}_0, \hat{\beta}_1)$ under Gaussian calibration errors. Then the distribution of $T_4(S_1^2, \ldots, S_k^2)$ is identical to its normal-theory form, independently of $(\hat{\beta}_0, \hat{\beta}_1)$ and the empirical calibration line.*

Proof The proof for Theorem 4(v) asserts $f(\boldsymbol{u}; \nu_1, \ldots, \nu_k) = \prod_{i=1}^k g_0(u_i; \nu_i/2, 2)$ as the conditional density for $\mathcal{L}(\nu_1 S_1^2/\hat{\beta}_1\omega_1^2, \ldots, \nu_k S_k^2/\hat{\beta}_1\omega_k^2 \mid \hat{\beta}_1)$, with $\nu_i = n_i - 1$ and $g_0(u; \alpha, \beta) = u^{\alpha-1}\text{e}^{-u/\beta}/\beta^\alpha\Gamma(\alpha)$, so that $\{S_1^2, \ldots, S_k^2\}$ are conditionally independent given $\hat{\beta}_1$. But since $T_4(S_1^2, \ldots, S_k^2)$ is scale-invariant, $\mathcal{L}[T_4(S_1^2, \ldots, S_k^2) \mid \hat{\beta}_1] = \mathcal{L}[T_4(S_1^2, \ldots, S_k^2)]$ unconditionally, to complete our proof. ■

It deserves note that meaningful comparisons among variances are necessarily scale-invariant. Moreover, it is seen that procedures based on $\{S_1^2, \ldots, S_k^2\}$ support tests for conditional hypotheses that $\{\text{var}(Y_{ij} \mid \hat{\beta}_1) = \hat{\beta}_1^2\omega^2 + \sigma_0^2; 1 \le i \le k\}$, or equivalently, $H_0 : \omega_1^2 = \omega_2^2 = \cdots = \omega_k^2$, to be tested against alternatives as appropriate. Theorem 5 applies in the case of both null and non–null distributions of invariant test statistics. Tests in common usage include

- modifications of Bartlett's [4] likelihood ratio test,
- Cochran's [5] test based on $S_{\max}^2/(S_1^2 + \cdots + S_k^2)$,
- Hartley's [6] F-max test based on the maximal ratio $\max\{S_i^2/S_j^2\}$, and
- Gnanadesikan's [7] simultaneous comparisons of treatment variances with a control.

To examine effects of calibration errors on the one-way analysis of variance for comparing means, we proceed conditionally given $\hat{\beta}_1$, first assuming that $\{\text{var}(Y_{ij} \mid \hat{\beta}_1) = \hat{\beta}_1^2\omega^2 + \sigma_0^2; 1 \le j \le n_i, 1 \le i \le k\}$, so that $V(\boldsymbol{Y} \mid \hat{\beta}_1) = \hat{\beta}_1^2\omega^2\boldsymbol{I}_n + \sigma_0^2\mathbf{1}_n\mathbf{1}'_n = \boldsymbol{\Xi}(\hat{\beta}_1)$ in the notation of Section 3.2. We are concerned with comparative inferences regarding elements of $\boldsymbol{\mu}(\beta_1) = \beta_1[\mu_1, \ldots, \mu_k]'$

from $E(\boldsymbol{Y}) = \beta_0 \mathbf{1}_n + \beta_1[\mu_1 \mathbf{1}'_{n_1}, \ldots, \mu_k \mathbf{1}'_{n_k}]'$. Recall that $\boldsymbol{I}_n = \boldsymbol{A}_0 + \boldsymbol{A}_1 + \boldsymbol{A}_2$ partitions $\boldsymbol{Y}'\boldsymbol{I}_n\boldsymbol{Y} = \boldsymbol{Y}'\boldsymbol{A}_0\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y}$ such that $\boldsymbol{Y}'\boldsymbol{A}_0\boldsymbol{Y} = n\overline{Y}^2$, with $\overline{Y}$ as the grand mean and $\boldsymbol{A}_0 = n^{-1}\mathbf{1}_n\mathbf{1}'_n$; $\boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y} = \sum_{i=1}^k n_i(\overline{Y}_i - \overline{Y})^2$; and $\boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y} = \sum_{i=1}^k \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$. To validate the Fisher–Cochran theorem conditionally requires that $\{\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta})\boldsymbol{A}_j = \mathbf{0}; i \neq j\}$. Moreover, scale parameters associated with the quadratic forms are found as $\{\kappa_i^2\boldsymbol{A}_i = \boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_i; i = 1, 2, 3\}$, whereas non-centrality parameters derive from expected mean squares. This program of study is carried out next in support of the following.

THEOREM 6 *Let* $\{Y_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Z_{ij}; 1 \leq j \leq n_i, 1 \leq i \leq k\}$ *be calibrated measurements in a one-way experiment such that* $\{(Z_{ij} - \mu_i)/\omega; 1 \leq j \leq n_i, 1 \leq i \leq k\}$ *are iid* $N_1(0, 1)$ *random variables independently of* $(\hat{\beta}_0, \hat{\beta}_1)$ *under Gaussian calibration errors.*

(i) *The analysis of variance test for equality of elements of* $\boldsymbol{\mu}(\beta_1) = \beta_1[\mu_1, \ldots, \mu_\kappa]'$, *pertaining to the group measurement means, is identical in level and power to its normal-theory form.*
(ii) *Supporting tests, based on linear contrasts among the group means, are identical in level and power to their normal-theory forms.*

Proof To validate the Fisher–Cochran theorem conditionally, observe $\{\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_0 = \mathbf{0}; i = 1, 2\}$, since $\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_0 = \mathbf{A}_i\big(\hat{\beta}_1^2\omega^2\boldsymbol{I}_n + \sigma_0^2\mathbf{1}_n\mathbf{1}'_n\big)\boldsymbol{A}_0$ and $\{\boldsymbol{A}_i\boldsymbol{A}_0 = \mathbf{0}; i = 1, 2\}$ Similarly $\{\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_j = \mathbf{0}; (i, j) = 1, 2, i \neq j\}$ since $\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_j = \boldsymbol{A}_i(\hat{\beta}_1^2\omega^2\boldsymbol{I}_n + n\sigma_0^2\boldsymbol{A}_0)\boldsymbol{A}_j$ and $\{\boldsymbol{A}_i\boldsymbol{A}_j = \mathbf{0}; (i, j) = 0, 1, 2, i \neq j\}$ from standard properties of the one-way classification. Scale parameters, as determined from $\big\{\kappa_i^2\boldsymbol{A}_i = \boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1)\boldsymbol{A}_i; i = 1, 2\big\}$ are found to be equal – namely $\big\{\kappa_i^2\boldsymbol{A}_i = \boldsymbol{A}_i\big(\hat{\beta}_1^2\omega^2\boldsymbol{I}_n + \sigma_0^2\mathbf{1}_n\mathbf{1}'_n\big)\boldsymbol{A}_i = \hat{\beta}_1^2\omega^2\boldsymbol{A}_i; i = 1, 2\big\}$ from idempotency together with the annihilation $\{\boldsymbol{A}_i\boldsymbol{I}_n = \mathbf{0}; i = 1, 2\}$, so that $\big\{\kappa_i^2 = \kappa^2; i = 1, 2\big\}$. Finally, the noncentrality parameters and degrees of freedom associated with $\{\boldsymbol{Y}'\boldsymbol{A}_i\boldsymbol{Y}; i = 1, 2\}$ are determined from their expected mean squares. These are $\{E(\boldsymbol{Y}'\boldsymbol{A}_i\boldsymbol{Y}|\hat{\beta}_1) = \mathrm{tr}(\boldsymbol{A}_i\boldsymbol{\Xi}(\hat{\beta}_1) + [\boldsymbol{\mu}(\hat{\beta}_1)]'\boldsymbol{A}_i\boldsymbol{\mu}(\hat{\beta}_1); i = 1, 2\}$. It follows directly that

$$E(\boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y}|\hat{\beta}_1) = \mathrm{tr}(\boldsymbol{A}_1\boldsymbol{\Xi}(\hat{\beta}_1)) + [\boldsymbol{\mu}(\hat{\beta}_1)]'\boldsymbol{A}_1\boldsymbol{\mu}(\hat{\beta}_1) = (k-1)\kappa^2 + \hat{\beta}_1^2\sum_{i=1}^{\kappa} n_i(\mu_i - \overline{\mu})^2$$

with $\overline{\mu} = \sum_{i=1}^k n_i\mu_i/n$. Similarly $E(\boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y}|\hat{\beta}_1) = \mathrm{tr}(\boldsymbol{A}_2\boldsymbol{\Xi}(\hat{\beta}_1)) + [\boldsymbol{\mu}(\hat{\beta}_1)]'\boldsymbol{A}_2\boldsymbol{\mu}(\hat{\beta}_1) = (n-k)\kappa^2$ since $[\boldsymbol{\mu}(\hat{\beta}_1)]'\boldsymbol{A}_2\boldsymbol{\mu}(\hat{\beta}_1) = \hat{\beta}_1^2\sum_{i=1}^k\sum_{j=1}^{n_i}(\mu_i - \mu_i)^2 = 0$. From these developments, we infer that the distribution of the ratio $F = (n-k)\boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y}/(k-1)\boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y}$ satisfies $\mathcal{L}(F|\hat{\beta}_1) = F(k-1, n-k, \lambda(\hat{\beta}_1))$ with $\lambda(\hat{\beta}_1) = \hat{\beta}_1^2\sum_{i=1}^k n_i(\mu_i - \mu)^2/\hat{\beta}_1^2\omega^2 = \sum_{i=1}^k n_i(\mu_i - \mu)^2/\omega^2$. Thus the conditional and unconditional distributions are identical, i.e., $\mathcal{L}(F|\hat{\beta}_1) = \mathcal{L}(F) = F(k-1, n-k, \lambda)$, with $\lambda = \sum_{i=1}^k n_i(\mu_i - \mu)^2/\omega^2$. ■

5. Diagnostics

5.1. *Objectives*

Calibration errors exact profound disturbances, both in models and in data–analytic procedures, as shown. Myriad calibrated data sets have been analysed to date, supported of late by an evolving battery of diagnostic tools. On these grounds, it is tempting to dismiss the present study as academic: in fact, these issues long since would have surfaced in practice, to be addressed accordingly. At issue is the capacity of known diagnostics to uncover calibration-induced irregularities as documented here. We now address these concerns with regard to induced correlations,

non-normality, mixture distributions having excessive tails, and possible outliers. For definiteness, we return to the case of a single sample as in Section 4.1.

5.2. *Correlation*

Neither the conditional $\left(\sigma_0^2\left(\hat{\beta}_1^2\sigma_Z^2+\sigma_0^2\right)\right)$ nor the unconditional $\left(\left(\sigma_0^2+\sigma_1^2\mu_Z^2\right)/\left(\kappa_2\sigma_Z^2+\sigma_0^2+\sigma_1^2\mu_Z^2\right)\right)$ correlations need to be negligible. Tests for correlation entail dispersion matrices $V(\boldsymbol{Y})=\tau^2\boldsymbol{\Xi}(\omega)$, for which $\boldsymbol{\Xi}(\omega)=(\boldsymbol{I}_n+\omega\boldsymbol{A})$ with $\boldsymbol{A}$ fixed and $\boldsymbol{\Xi}(\omega)\in\mathbb{S}_n^+$. Specializing gives $\tau^2\boldsymbol{\Xi}(\omega)$ as $\boldsymbol{\Sigma}(\rho)$ under the equicorrelation models encountered here. Tests of note are due to Durbin and Watson [8–10], Anderson and Anderson [11], Theil [12], and others, all based on versions of von Neumann's [13] ratio $U=\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}/\boldsymbol{R}'\boldsymbol{R}$, with $\boldsymbol{R}$ as the observed residuals and with $\boldsymbol{B}(n\times n)$ fixed; for further details, see [14]. However, here the unconditional distributions are all identical to their normal-theory forms as if $\mathcal{L}(\boldsymbol{R})=N_n(\boldsymbol{0},\sigma^2\boldsymbol{B}_n)$. This is seen from the proof for Theorem 3(ii), where $\mathcal{L}(\boldsymbol{R}|\hat{\beta}_1)=N_n(\boldsymbol{0},\hat{\beta}_1^2\sigma_Z^2\boldsymbol{B}_n)$ together with the scale invariance of $U=\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}/\boldsymbol{R}'\boldsymbol{R}$, assuring that $\mathcal{L}(\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}/\boldsymbol{R}'R|\hat{\beta}_1)=\mathcal{L}(\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}/\boldsymbol{R}'\boldsymbol{R})$ unconditionally. All such diagnostics for correlative dependencies are totally blind, both to the conditional $[V(\boldsymbol{Y}|\hat{\beta}_1)=\hat{\beta}_1^2\sigma_Z^2\boldsymbol{I}_n+\sigma_0^2\boldsymbol{1}_n\boldsymbol{1}_n']$ and unconditional $[V(\boldsymbol{Y})=\kappa_2\sigma_Z^2\boldsymbol{I}_n+(\sigma_0^2+\sigma_1^2\mu_Z^2)\boldsymbol{1}_n\boldsymbol{1}_n']$ dispersion structures. In short, demonstrated calibration-induced correlations cannot be discerned through conventional diagnostic tools.

5.3. *Non-normality*

Diagnostics for normality encompass both graphical and hypothesis testing procedures. Graphs include plots of ordered residuals against their normal-theory expectations. Common usage includes the scaled residuals $\{R_i/S_Y;\,1\le i\le n\}$ or the Studentized residuals $\{W_iR_i/S_Y;\,i=1,2,\ldots,n\}$, standardized so that $\mathrm{var}(W_iR_i)=\sigma_Y^2$; see Sections 2.12 and 5.7 of ref. [15], for example. In calibrated data, these residual plots are indistinguishable from those for the conventional Gaussian model $N_n(\mu\boldsymbol{1}_n,\sigma^2\boldsymbol{I}_n)$, whatever be the joint mixture density at Equation (1) for the calibrated measurements. This follows since $\mathcal{L}(\boldsymbol{R}/(\boldsymbol{R}'\boldsymbol{R})^{1/2}|\hat{\beta}_1)=\mathcal{L}(\boldsymbol{R}/(\boldsymbol{R}'\boldsymbol{R})^{1/2})$ from scale invariance, the latter as a scaled singular multivariate Student's t distribution having $\nu=n-1$ degrees of freedom, depending on neither $\hat{\beta}_1$ nor σ_Y^2.

Tests for normality include the regression tests of Shapiro and Wilk [16], known to be powerful against a wide range of alternatives, including skewed or distributions having short or very long tails, even in small samples; see [17], for example. These tests utilize statistics $W=\left(\sum_{i=1}^n w_iY_{[i]}\right)^2/(n-1)S_Y^2$, where $\{Y_{[1]}\le Y_{[2]}\le\cdots\le Y_{[n]}\}$ are the ordered values of $\{Y_1,\ldots,Y_n\}$, and $\{w_1,\ldots,w_n\}$ are fixed weights. Such tests would appear promising for detecting the non-standard mixture distributions of calibrated measurements, where

$$W=\frac{\left(\sum_{i=1}^n w_iY_{[i]}\right)^2}{(n-1)S_Y^2}=\frac{\left(\hat{\beta}_0\sum_{i=1}^n w_i+\hat{\beta}_1^2\sum_{i=1}^n w_iZ_{[i]}\right)^2}{(n-1)S_Y^2}. \tag{8}$$

However, since $\sum_{i=1}^n w_i=0$ for these tests, together with the identity $S_Y^2=\hat{\beta}_1^2S_Z^2$, it follows that $W=\left(\sum_{i=1}^n w_iZ_{[i]}\right)^2/(n-1)S_Z^2$. Then $\mathcal{L}(W|\hat{\beta}_1)=\mathcal{L}(W)$ holds unconditionally from cancellation. Briefly, all such regression tests fail to distinguish between Gaussian distributions, and the Gaussian mixtures of type (1). With regard to further variations on regression tests, as in [18], similar arguments show that none is able to distinguish between Gaussian distributions and their mixtures from calibrated measurements. Given the sample moments $\{m_r=\sum_{i=1}^n(Y_i-\overline{Y})^r;\,r=2,3,4\}$, tests based on the moment ratios $\left\{b_1=m_3^2/m_2^3,\,b_2=m_4/m_2^2\right\}$ are useful against skewed alternatives or distributions having excessive or short tails [18]. It is readily shown that these

ratios are precisely those obtainable from $\{Z_1, \ldots, Z_n\}$, so that their null distributions are identical to those for which $\mathcal{L}(\boldsymbol{Y}) = N_n(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{I}_n)$, whatever be the joint mixture distribution as in Equation (1). On the other hand, the preceding tests do offer a clear check on normality of the distribution of $\{Z_1, \ldots, Z_n\}$, on which the mixtures (1) are predicated.

Briefly, conventional Gaussian diagnostics are bereft of any capacity to distinguish between Gaussian errors, and Gaussian mixtures of type (1). Thus radical calibration-induced departures from Gaussian models cannot be discerned through routine screening using any of these diagnostics.

5.4. *Outliers*

Commonly used diagnostics for a shift in location or scale at observation Y_i include the studentized residuals $t_i = R_i / S_Y \sqrt{(1 - 1/n)}$ and the *R-Student* deletion diagnostic $Rt_i = R_i / S_{-i} \sqrt{(1 - 1/n)}$, where S_{-i} is the sample standard deviation found on deleting Y_i from $\{Y_1, \ldots, Y_n\}$. As mixture distributions may have heavy tails, and since conventional diagnostics for normality have failed, it is natural to ask whether outlier diagnostics might be sensitive to observations from mixtures of type (1). If so, then evidence for apparent outliers in calibrated data instead might be attributable to the calibration process itself. However, these diagnostics are all scale-invariant functions of the observed residuals $\{R_1, \ldots, R_n\}$, so that they are indistinguishable from the statistics derived from the standard Gaussian model $N_n(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{I}_n)$. Briefly, conventional outlier diagnostics cannot distinguish between Gaussian errors, and between heavy-tailed mixtures as in Equation (1), even if a shift in location or scale has occurred at observation Y_i.

Section 5 has re-examined whether conventional diagnostics can detect calibration-induced anomalies, including correlations, non-normality, distributions having excessive tails, and possible outliers. Even radical departures from conventional assumptions cannot be discerned through routine screening using any of the above-mentioned diagnostics. In summary, the present study cannot be dismissed as merely academic, as evidence for anomalies traceable to calibration could not have surfaced in practice through a battery of diagnostic tools as it has evolved to date.

6. Case studies

We apply the results of Section 4.1 to a numerical data set under the assumptions of Theorem 3. Table 1 gives the percent of purity (X) and the octane number (U) from a sample of $n = 11$ different gasoline production runs. Percent purity is determined readily, whereas octane numbers require expensive and time-consuming dynamic laboratory tests; hence the need for calibration.

The least-squares fit for $U = \beta_0 + \beta_1(X - \overline{X}) + \varepsilon$ has $\{n = 11, \hat{\beta}_0 = 87.2818, \hat{\sigma}_0 = 0.1846, \hat{\beta}_1 = 1.8546, \hat{\sigma}_1 = 0.5837\}$. Suppose that subsequent determinations of percent purity satisfy $\{\mathcal{L}(Z_i) = N_1(0, 1); 1 \le i \le n\}$, so that calibrated measurements are recovered as $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \le i \le n\}$ in units of octane number. Then the distribution of $\overline{Y}$ is the mixture of a normal distribution $N_1(\mu(t), \Xi(t))$, with $\mu(t) = \beta_0 + t\mu_Z$ and $\Xi(t) = t^2\sigma_Z^2/n + \sigma_0^2$, having the density $g_1(u; \mu(t), \Xi(t))$, and a mixing distribution $N_1(\beta_1, \sigma_1^2)$ having the density $\mathrm{d}G_1(t)$. For convenience, we write this as $\mathcal{L}(\overline{Y}) = N_1(\beta_0 + t\mu_Z, t^2\sigma_Z^2/n + \sigma_0^2) \Lambda_t N_1(\beta_1, \sigma_1^2)$, where Λ_t

Table 1. Percent of purity (X) and octane number (U) of gasoline.

X	99.8	99.7	99.6	99.5	99.4	99.3	99.2	99.1	99.0	98.9	98.8
U	88.6	86.4	87.2	88.4	87.2	86.8	86.1	87.3	86.4	86.6	87.1

designates the mixing operation. Accordingly, the density of $\overline{Y}$ is

$$f_1(u) = \int_{-\infty}^{\infty} g_1\left(u; \beta_0 + t\mu_Z, \frac{t^2\sigma_Z^2}{n} + \sigma_0^2\right) \mathrm{d}G_1(t)$$
$$= \int_{-\infty}^{\infty} \frac{\exp(-[u-(\beta_0+t\mu_Z)]^2/(2(t^2\sigma_Z^2/n+\sigma_0^2)))}{\sqrt{2\pi(t^2\sigma_z^2/n+\sigma_0^2)}} \frac{\exp(-((t-\beta_1)^2)/2\sigma_1^2)}{\sqrt{2\pi\sigma_1^2}} \mathrm{d}t, \quad (9)$$

a function of the parameters $\Omega = \{n, \beta_0, \sigma_0, \mu_Z, \sigma_Z, \beta_1, \sigma_1\}$, with skewness 0.1464×10^{-6} and kurtosis 3.855, and with conditional mean $E(\overline{Y}) = 87.2818$, given the empirical calibration.

Using Equation (9), we compute the 95% probability region for $\overline{Y}$ as (86.037, 88.526) compared with $\hat{\beta}_0 \pm 1.96|\hat{\beta}_1|\sigma_Z/\sqrt{11} = (86.184, 88.376)$ if $\overline{Y}$ were normal. This latter interval is actually a 92.2% probability region. In addition, the density $f_1(u)$ is bell-shaped but is not normal. Table 2 gives its moments (mean, variance) and moment ratios (skewness (γ), kurtosis (κ)) for selected values of the parameters Ω.

The scaled sample variance is a mixture of a gamma distribution, $G_0(\cdot,\cdot)$, with mixing distribution $\mathrm{d}G_2(w)$ as a non-central chi-squared distribution, to give $\mathcal{L}((n-1)S_Y^2/\sigma_1^2\sigma_Z^2) = G_0((n-1)/2, 2t)\Lambda_t\chi_1^2(\lambda = \beta_1^2/\sigma_1^2)$ with $E(S_Y^2/\sigma_Z^2\sigma_1^2) = (1+\lambda)$, which for the octane data is the conditional value $E(S_Y^2)/\sigma_Z^2\sigma_1^2 = 11.095$, so that $E(S_Y^2) = 3.780$. The mixture distribution has the density

$$f_0(u) = \frac{u^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)} \int_0^{\infty} w^{-\nu/2}\mathrm{e}^{-u/2w}\mathrm{d}G_2(w)$$
$$= \frac{u^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)} \int_0^{\infty} w^{-\nu/2}\mathrm{e}^{-u/2w} \left[\frac{\mathrm{e}^{-\lambda/2-u/2}}{2^{1/2}} \sum_{j=0}^{\infty} \left(\frac{\lambda}{4}\right)^j \frac{u^{j-1/2}}{j!\Gamma(1/2+j)}\right] \mathrm{d}w \quad (10)$$

with $\nu = (n-1)$.

If Y were normal with $\hat{\beta}_1$ a constant, then a 95% probability region for S_Y^2 could be found from $P[\chi^2(10; 0.025) < (n-1)S_Y^2/\hat{\beta}_1^2\sigma_Z^2 < \chi^2(10; 0.975)]$ or equivalently $P[1.1167 < S_Y^2 < 7.0449]$, which actually is a 74% probability region when variation in $\hat{\beta}_1$ is taken into account. The correct probability region is found by numerically integrating Equation (10) to get

Table 2. The moments (mean, variance) and moment ratios (skewness (γ), kurtosis (κ)) of $\mathcal{L}(\overline{Y})$ for selected values of the parameters $\Omega = \{n, \beta_0, \sigma_0, \mu_Z, \sigma_Z, \beta_1, \sigma_1\}$.

n	β_0	σ_0	μ_Z	σ_Z	β_1	σ_1	$E(\overline{Y})$	$\mathrm{var}(\overline{Y})$	γ	κ
10	1	1	1	1	1	1	2.0000	2.2000	0.1839	3.2851
20	1	1	1	1	1	1	2.0000	2.1000	0.0986	3.1463
20	0.5	1	1	1	1	1	1.5000	2.1000	0.0986	3.1463
20	2	1	1	1	1	1	3.0000	2.1000	0.0986	3.1463
20	1	0.5	1	1	1	1	2.0000	1.3500	0.1913	3.3539
20	1	2	1	1	1	1	2.0000	5.1000	0.0260	3.0248
20	1	1	0.5	1	1	1	1.5000	1.3500	0.0956	3.1070
20	1	1	2	1	1	1	3.0000	5.1000	0.0521	3.0940
20	1	1	1	0.5	1	1	2.0000	2.0250	0.0260	3.0373
20	1	1	1	2	1	1	2.0000	2.4000	0.3327	3.5417
20	1	1	1	1	0.5	1	1.5000	2.0625	0.0506	3.1463
20	1	1	1	1	2	1	3.0000	2.2500	0.1778	3.1452
20	1	1	1	1	1	0.5	2.0000	1.3125	0.0499	3.0267
20	1	1	1	1	1	2	2.0000	5.2500	0.0998	3.3614
10	1	0.5	1	2	1	2	2.0000	6.2500	0.6144	5.5559

$P[10.8 < (n-1)S_Y^2/\sigma_1^2\sigma_Z^2 < 336.5] = 0.95 = P[0.3680 < S_Y^2 < 11.46]$. We find that using the first 20 terms in the infinite sum is adequate.

For the density of $t_0^2 = n(\overline{Y} - \mu_Y^0)^2/S_Y^2$, set

$$\left\{ \nu = n-1, \delta = \frac{(\mu_Y - \mu_Y^0)^2}{\sigma_1^2\sigma_Z^2}, \delta(\hat{\beta}_1) = \frac{\delta}{\hat{\beta}_1^2/\sigma_1^2} \right\}.$$

Its density is found on mixing the non-central $t^2(\nu, \delta/t)$ over a non-central chi-squared as the mixing distribution, which we write as $\mathcal{L}(t_0^2) = t^2(\nu, \delta/t)\Lambda_t\chi_1^2(\lambda = \beta_1^2/\sigma_1^2)$. The density is given by

$$g(u;\nu,\delta,\lambda) = \int_0^\infty \frac{1}{\nu} \sum_{j=0}^\infty \frac{(\delta/t/2)^j e^{(-\delta/t/2)}(u/\nu)^{-1/2+j}}{j!B((1+2j)/2, \nu/2)(1+u/\nu)^{1/2+\nu/2+j}} \frac{e^{-\lambda/2-t/2}}{2^{1/2}}$$
$$\times \sum_{k=0}^\infty \frac{(\lambda/4)^k t^{k-1/2}}{k!\Gamma((1+2k)/2)} dt. \tag{11}$$

For the first sum, we use the first $N_1 = 15$ terms, and for the second sum the first $N_2 = 30$ terms.

This distribution is useful for computing the power of the t^2 test. For example, with $n = 11$ so $\nu = 10$, the 95% critical value is 4.9646 with $\delta = 0$. Table 3 gives the power of the test for $\delta = \{0, 1, 4, 9\}$ and $\lambda = \{1, 4, 9\}$

For the octane data, the power of the test of $H_0 : \mu_Y = \mu_Y^0$, with $(\mu_Y - \mu_Y^0)^2 = 1$, has $\nu = 10$, $\delta = (\mu_Y - \mu_Y^0)^2/\sigma_Z^2\sigma_1^2 = [(1)(0.5837)]^{-2} = 2.9351$, $\lambda = \beta_1^2/\sigma_1^2 = (1.8546/0.5837)^2 = 10.0953$. The power of the test is 90%.

An equivalent form for $\mathcal{L}(t_0^2)$ is based on t_0 as $\mathcal{L}(t_0) = t(\nu, \delta_0/s)\Lambda_s\sqrt{(\chi_1^2(\lambda = \beta_1^2/\sigma_1^2))}$, mixing over a shifted half-normal distribution. Its density is

$$f_{t_0}(u) = \int_0^\infty \frac{e^{(-(\delta_0/s)^2)/2}\Gamma((\nu+1)/2)(\nu/(\nu+u^2))^{\nu/2+1/2}}{\sqrt{\pi\nu}(\Gamma/2)}$$
$$\times \sum_{j=0}^\infty \frac{\Gamma((\nu+j+1)/2)}{j!\Gamma((\nu+1)/2)} \left(\frac{\sqrt{2}u\delta_0/s}{\sqrt{\nu+u^2}} \right)^j d\sqrt{G_2}(s) \tag{12}$$

having the non-centrality parameter $\delta_0 = \sqrt{\delta}$ and

$$d\sqrt{G_2}(s) = \frac{e^{-(s-\lambda_0)^2/2} + e^{-(-s-\lambda_0)^2/2}}{\sqrt{2\pi}} ds$$

with the non-centrality parameter $\lambda_0 = \sqrt{\lambda}$. This series has faster convergence and we used $N = 20$ terms in the preceding power calculations with non-centrality parameters $\delta_0 = \sqrt{\delta}$ and

Table 3. Power for the test $H_0 : \mu_Y = \mu_Y^0$, against $H_1 : \mu_Y \neq \mu_Y^0$ for $\delta \in \{0, 1, 4, 9\}$ and $\lambda \in \{1, 4, 9\}$.

	λ		
δ	1	4	9
0	0.950	0.950	0.950
1	0.691	0.863	0.928
4	0.485	0.742	0.876
9	0.329	0.608	0.799

$\lambda_0 = \sqrt{\lambda}$. Computations reported here were executed by the second author using the Maple software package.

7. Conclusions

In summary, the widespread and necessary use of calibration may have devastating effects, even on elementary data-analytic procedures pertaining to location and scale parameters. It is unfortunate that these difficulties cannot be flagged by the ever expanding use of available diagnostic tools. It is thus incumbent on knowledgeable users of statistical methodology, and the statistical consultants advising them, to assess the extent of these difficulties as they might impact the analysis and interpretation of data in a particular experimental setting. Let the user be forewarned. Fortunately, comparisons among means and among variances, in the context of comparative one-way experiments, are largely unaffected by the use of calibrated instruments when subject to errors of calibration, provided that the results are interpreted accordingly.

References

[1] C. Loewner, *ÜUber monotone Matrixfunktionen*, Math Z. 38 (1934), pp. 177–216.
[2] S. Sherman, *A theorem on convex sets with applications*, Ann. Math. Statist. 25 (1955), pp. 763–766.
[3] R.G. Krutchkoff, *The calibration problem and closeness*, J. Statist. Comput. Simul. 1 (1971), pp. 87–95.
[4] M.S. Bartlett, *Properties of sufficiency and statistical tests*, Proc. Royal Soc. A 160 (1937), pp. 268–282.
[5] W.G. Cochran, *The distribution of the largest of a set of estimated variances as a fraction of their total*, Ann. Eugenics 11 (1941), pp. 47–52.
[6] H.O. Hartley, *The maximum F-ratio as short-cut test for heterogeneity of variances*, Biometrika 37 (1950), pp. 308–312.
[7] R. Gnanadesikan, *Equality of more than two variances and more than two dispersion matrices against certain alternatives*, Ann. Math. Statist. 30 (1959), pp. 177–184.
[8] J. Durbin and G.S. Watson, *Testing for serial correlation in least squares regression. I*, Biometrika 37 (1950), pp. 409–428.
[9] J. Durbin and G.S. Watson, *Testing for serial correlation in least squares regression. II*, Biometrika 38 (1951), pp. 159–178.
[10] J. Durbin and G.S. Watson, *Testing for serial correlation in least squares regression. III*, Biometrika 58 (1971), pp. 1–19.
[11] R.L. Anderson and T.W. Anderson, *Distribution of the circular serial correlation for residuals from a fitted Fourier series*, Ann. Math. Statist. 21 (1950), pp. 59–81.
[12] H. Theil, *The analysis of disturbances in regression analysis*, J. Am. Statist. Assoc. 60 (1965), pp. 1067–1079.
[13] J. von Neumann, *Distribution of mean square successive difference to the variance*, Ann. Math. Statist. 12 (1941), pp. 367–395.
[14] T. Kariya, *A robustness property of the tests for serial correlation*, Ann. Statist. 5 (1977), pp. 1212–1220.
[15] R.H. Myers, *Classical and Modern Regression with Applications*, 2nd edn., PWS-Kent, Boston, (1990).
[16] S.S. Shapiro and M.B. Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika 52 (1965), pp. 591–611.
[17] J.P. Royston, *Shapiro–Wilk W statistics*, in *Encyclopedia of Statistical Sciences*, Vol. 8, S. Kotz, N.L. Johnson, and C.B. Read, eds., Wiley, New York, 1988, pp. 430–431.
[18] R.B. D'Agostino, Departures from normality, tests for, in *Encyclopedia of Statistical Sciences*, Vol. 2, S. Kotz, N.L. Johnson, and C.B. Read, eds., Wiley, New York, 1982, pp. 315–324.
[19] E.L. Lehmann, *Testing Statistical Hypotheses*, 2nd edn., John Wiley, New York, 1986.
[20] T.W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. 6 (1955), pp. 170–176.
[21] D.R. Jensen, *Ordering ellipsoidal measures: Scale and peakedness orderings*, SIAM J. Appl. Math. 44 (1984), pp. 1226–1231.

Appendix

It is germane to examine the comparative concentration of probability measures on $\mathbb{R}^n$. Following Sherman [2], the measure $\mu(\cdot)$ is said to be *more peaked* about $\mathbf{0} \in \mathbb{R}^n$ than $\nu(\cdot)$ if and only if $\mu(A) \geq \nu(A)$ for every set A in the class $\mathcal{C}(n)$ comprising the convex sets in $\mathbb{R}^n$ symmetric under reflection through $\mathbf{0} \in \mathbb{R}^n$. For scale mixtures of Gaussian measures on $\mathbb{R}^n$, their peakedness ordering is tantamount to the stochastic ordering of their mixing distributions. Details follow.

LEMMA A1 *Let* $\mathrm{GM}_n(\boldsymbol{\theta}, \boldsymbol{\Xi}, G_1)$ *and* $\mathrm{GM}_n(\boldsymbol{\theta}, \boldsymbol{\Xi}, G_2)$ *be Gaussian mixtures on* $\mathbb{R}_n$ *of type* (2) *having mixing distributions* $G_1(\cdot)$ *and* $G_2(\cdot)$ *on* $\mathbb{R}^1_+$. *Then* $\mathrm{GM}_n(\boldsymbol{\theta}, \boldsymbol{\Xi}, G_1)$ *is more peaked about* $\boldsymbol{\theta} \in \mathbb{R}^n$ *than* $\mathrm{GM}_n(\boldsymbol{\theta}, \boldsymbol{\Xi} G_2)$ *if and only if* $G_1(t) \leq G_2(t)$ *for every* $t > 0$.

Proof The ordering $G_1(t) \leq G_2(t)$; i.e., that $G_1(\cdot)$ is stochastically larger than $G_2(\cdot)$, holds if and only if there are increasing functions $\{\psi_1(\cdot), \psi_2(\cdot)\}$ ordered pointwise as $\psi_1(t) \geq \psi_2(t)$, together with a random variable U, such that $G_1(t) = P(\psi_1(U) \leq t)$ and $G_2(t) = P(\psi_2(U) \leq t)$; see, for example, [19, Lemma 1, p. 84]. Accordingly, we provisionally write $\mu(A) = \int_A f(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\Xi}, G_1)\mathrm{d}\boldsymbol{x}$ and $\nu(A) = \int_A f(\boldsymbol{x}; \boldsymbol{\theta}\boldsymbol{\Xi}, G_2)\mathrm{d}\boldsymbol{x}$ and their difference as

$$\mu(A) - \nu(A) = \int_0^\infty \int_A [g(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\Xi}/\psi_1(t)) - g(\boldsymbol{x}; \boldsymbol{\theta}\boldsymbol{\Xi}/\psi_2(t))]\mathrm{d}\boldsymbol{x}\,\mathrm{d}G(t).$$

Given that $G_1(t) \leq G_2(t)$, so that $\psi_1(t) \geq \psi_2(t)$, the ordering $\int_A [g(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\Xi}/\psi_1(t)) - g(\boldsymbol{x}; \boldsymbol{\theta}\boldsymbol{\Xi}/\psi_2(t))]\mathrm{d}\boldsymbol{x} \geq 0$ follows point wise for each fixed $t \in \mathbb{R}^1_+$ from [20, Corollary 3], since $\boldsymbol{\Xi}/\psi_2(t) \succeq_L \boldsymbol{\Xi}/\psi_1(t)$ uniformly in t. That $[\mu(A) - \nu(A)] \geq 0$ now follows directly. Conversely, suppose that $\mu(A) \geq \nu(A)$. We now apply the converse to Anderson's [20, Corollary 3], as proved in [21], to infer that $\psi_1(t) \geq \psi_2(t)$ for each $t > 0$, thus establishing the necessity of the condition $G_1(t) \leq G_2(t)$, to complete our proof. ■