axioms

MDPI

Article

On the Structural Distortion Induced by the Inverse Box–Cox Transformation

Rui Gonçalves

Department of Civil and Georesources Engineering, Faculty of Engineering, Universidade do Porto, 4200-465 Porto, Portugal; rjasg@fe.up.pt

Abstract

The Box–Cox transformation is widely used to improve normality, stabilize variance, and enable Gaussian-based modelling in a transformed scale. After model fitting, conditional summaries are often mapped back to the original scale by applying the inverse transformation. This paper shows that this transform–fit–inverse procedure has a structural limitation: nonlinear inverse transformations do not, in general, preserve conditional expectations. Equivalently, conditional expectation and nonlinear inversion do not commute. Within the Box–Cox Gaussian framework, the admissible domain of the inverse transformation leads naturally to a truncated normal formulation in the transformed scale. Under this formulation, we derive a second-order decomposition showing that the original-scale conditional mean differs from the inverse-transformed truncated conditional mean by a curvature-driven correction term depending on the truncated conditional variance. The usual untruncated Gaussian expression is recovered as a local approximation when the inadmissible probability is negligible. A numerical sensitivity analysis, focused on $0 \leq \lambda_Y \leq 1$, illustrates how the distortion depends on the transformation parameter, correlation, and conditional dispersion. A real-data illustration using medical insurance charges further shows that the discrepancy can be visible in an applied regression setting and is not removed by changing the transformation of the explanatory variable. The results distinguish this structural invariance problem from classical retransformation bias and show that inverse-transformed fitted curves should be interpreted as transformation-induced structural curves, not automatically as conditional mean functions on the original scale.

Keywords: power normal distribution; Box–Cox transformation; conditional expectation; regression distortion; nonlinear transformation; latent Gaussian structure

MSC: 62J05; 62E10; 62H12; 60E05

Academic Editor: Xuejun Wang

Received: 29 May 2026
Revised: 2 July 2026
Accepted: 6 July 2026
Published: 10 July 2026

1. Introduction

Transformation-based modelling is a standard strategy in statistics. A nonlinear transformation is often applied to the response variable, or to several variables, in order to improve normality, stabilize variance, or simplify the dependence structure. After fitting a model in the transformed scale, fitted values, regression curves, or conditional summaries are commonly mapped back to the original scale by applying the inverse transformation. This transform–fit–inverse procedure is natural and widely used, but it raises a fundamental question: which statistical structures are preserved when a model fitted in the transformed scale is returned to the original scale?

This question is particularly important for conditional expectations. In regression analysis, the conditional mean is often the primary object of interpretation and prediction. However, conditional expectation is a linear operator, whereas the inverse of a nonlinear transformation is itself nonlinear. As a consequence, the inverse transformation of a conditional mean in the transformed scale does not generally coincide with the conditional mean in the original scale. In operator terms, nonlinear transformations do not commute with conditional expectation.

The Box–Cox transformation, introduced by [1], provides a classical and analytically tractable setting in which this issue can be studied. It is widely used for variance stabilization, improvement of normality, and simplification of functional relationships in statistical modelling [2–4]. It is also closely related to the Power Normal distribution family introduced by [5], which provides a probabilistic framework for transformation-based Gaussian modelling [6–8]. The transformation continues to be used in modern applications, including data-driven modelling and machine-learning contexts [9,10].

The phenomenon studied here is related to, but distinct from, the classical retransformation bias problem. Retransformation bias is primarily concerned with corrected prediction or mean estimation on the original scale after fitting a model in a transformed scale [11–15]. In contrast, the present paper addresses a structural invariance question: whether the conditional mean operator itself is preserved by a transform–fit–inverse modelling pipeline.

The contribution of this paper is to make explicit this structural non-commutativity principle in the Box–Cox framework. If h is a nonlinear inverse transformation and V is a transformed response, then in general

$$\mathbb{E}[h(V) \mid U] \neq h\{\mathbb{E}[V \mid U]\}.$$

Equality can hold under restrictive conditions, such as affine transformations or degenerate conditional distributions, but it is not a generic property of nonlinear transformation-based models. Thus, the inverse image of a conditional mean in the transformed scale is generally a structural curve induced by the transformation, not the conditional mean function in the original scale.

Within the Box–Cox setting, the admissible-domain issue must also be taken into account. The inverse Box–Cox map is defined only on

$$\mathcal{D}_\lambda = \{u \in \mathbb{R} : 1 + \lambda u > 0\}.$$

Thus, an unrestricted Gaussian latent variable is not globally compatible with the inverse Box–Cox map whenever $\lambda \neq 0$. For this reason, the exact probabilistic formulation considered in this paper is based on the Gaussian density restricted to the admissible domain, that is, on a truncated normal formulation in the transformed scale. In that formulation, the discrepancy between the original-scale conditional mean and the inverse-transformed conditional mean is governed, to second order, by the curvature of the inverse transformation and by the conditional variance in the transformed domain.

Transformation methods have a long history in statistical modelling and have been used not only to improve approximate normality, but also to reduce skewness, stabilize variance and simplify regression [1,16]. The Box–Cox transformation is one of the most widely used members of this tradition, but it is part of a broader class of transformation-based methods, including later extensions such as the Yeo–Johnson transformation for variables that may take nonpositive values [17]. Multivariate versions of the Box–Cox transformation have also been proposed in order to address joint normality and dependence structure [18].

Related transformation ideas also appear in Gaussian copula modelling and in modern statistical learning procedures, where nonlinear marginal transformations are often used to separate marginal behaviour from dependence modelling. The present paper focuses on a different, but related, issue: the effect of nonlinear inverse transformations on conditional means after fitting a model in the transformed scale.

The transform–fit–invert pipeline considered in this paper can be summarized schematically as

$$(X, Y) \longrightarrow (U, V) = \{g_{\lambda_X}(X), g_{\lambda_Y}(Y)\} \longrightarrow \mathbb{E}[V \mid U = u] \longrightarrow g_{\lambda_Y}^{-1}\{\mathbb{E}[V \mid U = u]\}.$$

The structural discrepancy arises in the final step because

$$g_{\lambda_Y}^{-1}\{\mathbb{E}[V \mid U = u]\} \neq \mathbb{E}\{g_{\lambda_Y}^{-1}(V) \mid U = u\}$$

unless the inverse transformation is affine or the conditional dispersion is degenerate.

The paper is organized as follows. Section 2 formulates the general non-commutativity principle for conditional expectations under nonlinear transformations. Section 3 introduces the Box–Cox transformation, its inverse, and the corresponding Gaussian framework with admissible-domain truncation. Section 4 derives the second-order distortion formula. Section 5 presents a numerical sensitivity analysis. Section 6 provides a real-data illustration and a sensitivity analysis with respect to the transformation of the explanatory variable. Section 7 discusses practical implications and concludes.

2. Conditional Expectation, Projection, and Non-Commutativity

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $V \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. For a sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, the conditional expectation $\mathbb{E}[V \mid \mathcal{G}]$ is the orthogonal projection of V onto the closed subspace $L^2(\mathcal{G})$ of $\mathcal{G}$-measurable square-integrable random variables. Thus, if $P_{\mathcal{G}}$ denotes this projection operator, then

$$P_{\mathcal{G}} V = \mathbb{E}[V \mid \mathcal{G}].$$

In particular, conditional expectation is a linear operator.

In contrast, if $h : \mathbb{R} \to \mathbb{R}$ is nonlinear, the map $V \mapsto h(V)$ is nonlinear. Hence there is, in general, no reason to expect

$$P_{\mathcal{G}}\{h(V)\} = h(P_{\mathcal{G}} V),$$

or equivalently

$$\mathbb{E}[h(V) \mid \mathcal{G}] = h\{\mathbb{E}[V \mid \mathcal{G}]\}.$$

The failure of this identity is the non-commutativity mechanism studied in this paper.

Theorem 1. ***Non-commutativity*** *of nonlinear transformations and conditional expectation. Let $h : I \to \mathbb{R}$ be a continuous function on an interval $I \subseteq \mathbb{R}$. Suppose that, for every square-integrable random variable V taking values in I, and for every sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$,*

$$\mathbb{E}[h(V) \mid \mathcal{G}] = h\{\mathbb{E}[V \mid \mathcal{G}]\}.$$

Then h is affine on I. Equivalently, no genuinely nonlinear transformation can preserve conditional expectations for all admissible conditional distributions.

Proof. It is enough to consider the trivial σ-algebra $\mathcal{G} = \{\emptyset, \Omega\}$. Then the assumed identity becomes

$$\mathbb{E}[h(V)] = h\{\mathbb{E}[V]\}.$$

Let $a, b \in I$ and $p \in (0, 1)$. If V is such that

$$\mathbb{P}(V = a) = p, \qquad \mathbb{P}(V = b) = 1 - p,$$

then

$$\mathbb{E}[V] = pa + (1-p)b, \qquad \mathbb{E}[h(V)] = ph(a) + (1-p)h(b).$$

The assumed identity therefore implies

$$ph(a) + (1-p)h(b) = h\{pa + (1-p)b\}.$$

Thus h satisfies Jensen's equality on I. Since h is continuous, it follows that h is affine on I. $\square$

The continuity assumption is used only to exclude pathological solutions of Jensen's functional equation. The same conclusion holds under weaker regularity assumptions, such as measurability or local boundedness on an interval. Under such assumptions, Jensen's equality implies that h is affine. In the present statistical setting this regularity requirement is natural, since the transformations used in transformation-based modelling, including the Box–Cox transformation, are continuous and smooth on their admissible domains.

The affine case is the exceptional case. If $h(v) = a + bv$, then

$$\mathbb{E}[h(V) \mid \mathcal{G}] = a + b\mathbb{E}[V \mid \mathcal{G}] = h\{\mathbb{E}[V \mid \mathcal{G}]\}.$$

Thus affine transformations preserve conditional expectations. The Taylor expansions derived below do not establish the non-commutativity itself; rather, they quantify its leading contribution in smooth transformation models.

3. The Box–Cox Transformation and Gaussian Framework

For $\lambda \neq 0$, the Box–Cox transformation is defined by

$$g_\lambda(x) = \frac{x^\lambda - 1}{\lambda}, \qquad x > 0,$$

with inverse

$$h_\lambda(u) = (1 + \lambda u)^{1/\lambda}, \qquad 1 + \lambda u > 0.$$

For $\lambda = 0$, the transformation is understood in the limiting sense,

$$g_0(x) = \log x, \qquad h_0(u) = \exp(u).$$

The admissible domain of the inverse transformation is

$$\mathcal{D}_\lambda = \{u \in \mathbb{R} : 1 + \lambda u > 0\}.$$

Thus $\mathcal{D}_\lambda = (-1/\lambda, \infty)$ if $\lambda > 0$, $\mathcal{D}_\lambda = (-\infty, -1/\lambda)$ if $\lambda < 0$, and $\mathcal{D}_0 = \mathbb{R}$.

The first two derivatives of the inverse transformation are

$$h'_\lambda(u) = (1 + \lambda u)^{1/\lambda - 1}, \qquad h''_\lambda(u) = (1 - \lambda)(1 + \lambda u)^{1/\lambda - 2}.$$

Consequently, the inverse Box–Cox transformation is affine only when $\lambda = 1$, for which

$$g_1(x) = x - 1, \qquad h_1(u) = 1 + u, \qquad h_1''(u) = 0.$$

For every $\lambda \neq 1$, the inverse map has nonzero curvature on its admissible domain. Since the leading contribution to the discrepancy between

$$\mathbb{E}[h_\lambda(V) \mid U]$$

and

$$h_\lambda\{\mathbb{E}[V \mid U]\}$$

depends on the curvature of h_λ, the parameter λ controls not only the marginal transformation of the data but also the induced conditional-mean distortion.

We now specialize to a transformation-based Gaussian framework. Let $X > 0$ and $Y > 0$ be random variables on the original scale, and define

$$U = g_{\lambda_X}(X), \qquad V = g_{\lambda_Y}(Y),$$

where the transformation parameters may be different. A natural Gaussian reference model in the transformed scale is

$$(U, V) \sim \mathcal{N}_2(\mu, \Sigma),$$

with

$$\mu = \begin{pmatrix} \mu_U \\ \mu_V \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_U^2 & \rho\sigma_U\sigma_V \\ \rho\sigma_U\sigma_V & \sigma_V^2 \end{pmatrix},$$

where $\sigma_U > 0$, $\sigma_V > 0$, and $-1 < \rho < 1$.

Because the inverse Box–Cox maps are defined only on their admissible domains, the exact probabilistic formulation is the Gaussian density restricted to

$$\mathcal{A} = \mathcal{D}_{\lambda_X} \times \mathcal{D}_{\lambda_Y}.$$

Thus the transformed vector is interpreted as

$$(U, V) \mid (U, V) \in \mathcal{A}.$$

The unrestricted Gaussian distribution is used only as a reference density, or as an approximation when the probability of the inadmissible region is negligible. For a fixed admissible value of u, the Gaussian reference conditional distribution is

$$V \mid U = u \sim \mathcal{N}(m(u), \tau^2),$$

where

$$m(u) = \mu_V + \rho\frac{\sigma_V}{\sigma_U}(u - \mu_U), \qquad \tau^2 = \sigma_V^2(1 - \rho^2).$$

Here τ^2 denotes the conditional variance in the transformed scale, distinct from the marginal variance σ_V^2. In the exact admissible-domain formulation, this conditional normal distribution is truncated to $\mathcal{D}_{\lambda_Y}$. We denote the corresponding truncated conditional mean and variance by

$$m_T(u) = \mathbb{E}[V \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}]$$

and

$$\tau_T^2(u) = \mathrm{Var}(V \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}).$$

https://doi.org/10.3390/axioms15070519

When $\lambda_Y = 0$, no truncation is required and $m_T(u) = m(u)$, $\tau_T^2(u) = \tau^2$.

The structural discrepancy in the exact admissible-domain formulation is

$$\Delta_T(u) = \mathbb{E}[h_{\lambda_Y}(V) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] - h_{\lambda_Y}\{m_T(u)\}.$$

Since $X = h_{\lambda_X}(U)$, conditioning on $X = x$ is equivalent to conditioning on $U = g_{\lambda_X}(x)$. Thus,

$$\mathbb{E}[Y \mid X = x] = \mathbb{E}[h_{\lambda_Y}(V) \mid U = g_{\lambda_X}(x),\ V \in \mathcal{D}_{\lambda_Y}],$$

whereas the inverse-transformed truncated conditional mean is

$$h_{\lambda_Y}\big(m_T\{g_{\lambda_X}(x)\}\big).$$

4. Second-Order Distortion Formula

We now derive an explicit local representation of the structural distortion. In the exact admissible-domain formulation, we write

$$V = m_T(u) + \varepsilon_T,$$

where

$$\mathbb{E}[\varepsilon_T \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] = 0, \qquad \mathbb{E}[\varepsilon_T^2 \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] = \tau_T^2(u).$$

Since $Y = h_{\lambda_Y}(V)$, the original-scale conditional mean is

$$\mathbb{E}[Y \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] = \mathbb{E}[h_{\lambda_Y}(V) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}],$$

whereas the inverse-transformed truncated conditional mean is $h_{\lambda_Y}\{m_T(u)\}$. The following result should therefore be understood as a local small-dispersion expansion. More precisely, it describes the leading contribution to the discrepancy as the truncated conditional distribution of $V \mid U = u$ concentrates around its truncated conditional mean $m_T(u)$. In this sense, the remainder term is interpreted in the asymptotic regime $\tau_T^2(u) \to 0$. The second-order formula should be interpreted as a local small-dispersion approximation. Its accuracy depends on the concentration of the truncated conditional distribution of $V \mid U = u$ around $m_T(u)$ and on the smoothness of the inverse transformation over the region where this conditional distribution places most of its mass.

If h_{λ_Y} is three times continuously differentiable on an interval $I_u \subset \mathcal{D}_{\lambda_Y}$ containing the relevant support of the truncated conditional distribution, Taylor's theorem with remainder gives

$$h_{\lambda_Y}(V) = h_{\lambda_Y}\{m_T(u)\} + h'_{\lambda_Y}\{m_T(u)\}(V - m_T(u)) + \frac{1}{2}h''_{\lambda_Y}\{m_T(u)\}(V - m_T(u))^2 + R_3(u),$$

with

$$|R_3(u)| \le \frac{M_3(u)}{6}|V - m_T(u)|^3, \qquad M_3(u) = \sup_{z \in I_u}\left|h_{\lambda_Y}^{(3)}(z)\right|.$$

Taking conditional expectations gives

$$\left|\mathbb{E}[R_3(u) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}]\right| \le \frac{M_3(u)}{6}\mathbb{E}\Big[|V - m_T(u)|^3 \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}\Big].$$

For the Box–Cox inverse transformation,

$$h_\lambda^{(3)}(z) = (1-\lambda)(1-2\lambda)(1+\lambda z)^{1/\lambda - 3},$$

with the limiting exponential case obtained when $\lambda = 0$. Therefore, the approximation is expected to be most reliable when the conditional dispersion in the transformed scale is small and when $m_T(u)$ is not close to the boundary of the admissible domain $\mathcal{D}_{\lambda_Y}$.

Proposition 1. ***Second-order*** *structural distortion. Assume that $m_T(u) \in \mathcal{D}_{\lambda_Y}$ and that the truncated conditional distribution of $V \mid U = u$ is concentrated in a neighbourhood where h_{λ_Y} is twice differentiable. Then, in the local small-dispersion regime $\tau_T^2(u) \to 0$,*

$$\Delta_T(u) = \frac{1}{2}\tau_T^2(u) h''_{\lambda_Y}\{m_T(u)\} + o\Big(\tau_T^2(u)\Big),$$

where

$$\Delta_T(u) = \mathbb{E}[h_{\lambda_Y}(V) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] - h_{\lambda_Y}\{m_T(u)\}.$$

Proof. Using

$$V = m_T(u) + \varepsilon_T,$$

we expand h_{λ_Y} around $m_T(u)$:

$$h_{\lambda_Y}\{m_T(u) + \varepsilon_T\} = h_{\lambda_Y}\{m_T(u)\} + \varepsilon_T h'_{\lambda_Y}\{m_T(u)\} + \frac{\varepsilon_T^2}{2} h''_{\lambda_Y}\{m_T(u)\} + o(\varepsilon_T^2).$$

Taking conditional expectations under $U = u$ and $V \in \mathcal{D}_{\lambda_Y}$, and using the fact that $\mathbb{E}[\varepsilon_T \mid U = u, V \in \mathcal{D}_{\lambda_Y}] = 0$, gives

$$\mathbb{E}[h_{\lambda_Y}(V) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}] = h_{\lambda_Y}\{m_T(u)\} + \frac{1}{2} h''_{\lambda_Y}\{m_T(u)\}\tau_T^2(u) + o\Big(\tau_T^2(u)\Big).$$

□

Using

$$h''_{\lambda_Y}(z) = (1 - \lambda_Y)(1 + \lambda_Y z)^{1/\lambda_Y - 2},$$

the leading term becomes

$$\Delta_T(u) \approx \frac{1}{2}\tau_T^2(u)(1 - \lambda_Y)\{1 + \lambda_Y m_T(u)\}^{1/\lambda_Y - 2}.$$

Thus the distortion is governed by the transformed-scale conditional variance, the departure from the affine case, and the location of $m_T(u)$ within the admissible domain. Expressed as a function of the original covariate x, with $u = g_{\lambda_X}(x)$,

$$\Delta_{T,X}(x) \approx \frac{1}{2}\tau_T^2\{g_{\lambda_X}(x)\}(1 - \lambda_Y)\big[1 + \lambda_Y m_T\{g_{\lambda_X}(x)\}\big]^{1/\lambda_Y - 2}.$$

This shows that the discrepancy between the true conditional mean in the original scale and the inverse-transformed transformed-scale conditional mean is not a constant bias, but a structural function of the covariate.

When the probability assigned to the inadmissible region is negligible, the truncated conditional moments are close to their untruncated Gaussian counterparts:

$$m_T(u) \approx m(u), \qquad \tau_T^2(u) \approx \tau^2 = \sigma_V^2(1 - \rho^2).$$

In this local approximation,

$$\Delta(u) \approx \frac{1}{2}\sigma_V^2(1 - \rho^2)(1 - \lambda_Y)\{1 + \lambda_Y m(u)\}^{1/\lambda_Y - 2}.$$

The logarithmic case $\lambda_Y = 0$ provides a useful benchmark. Then $h_0(z) = \exp(z)$ and $h_0''(z) = \exp(z)$. Since $\mathcal{D}_0 = \mathbb{R}$, no truncation is required. If $V \mid U = u$ is Gaussian, the exact expression is

$$\mathbb{E}[\exp(V) \mid U = u] = \exp\left\{ m(u) + \frac{1}{2}\sigma_V^2(1-\rho^2) \right\}.$$

Hence

$$\Delta(u) = \exp\{m(u)\}\left[\exp\left\{\frac{1}{2}\sigma_V^2(1-\rho^2)\right\} - 1\right],$$

showing explicitly that the inverse-transformed conditional mean $\exp\{m(u)\}$ is not the conditional mean in the original scale.

Figure 1 illustrates the same mechanism numerically. The true conditional mean on the original scale is compared with the inverse-transformed linear predictor obtained from the transformed scale. Although the two curves are nearly indistinguishable over most of the domain, the zoom on the right tail reveals a small but systematic discrepancy. This confirms visually that the original-scale conditional mean is not, in general, simply the inverse image of the linear conditional mean in the transformed scale.

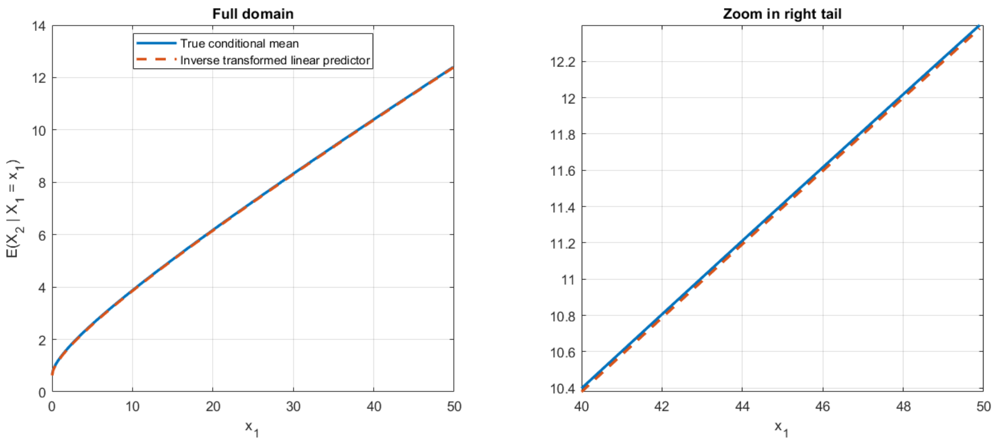

Figure 1. Comparison between the true conditional mean on the original scale and the inverse-transformed linear predictor obtained from the transformed scale. The curves are nearly indistinguishable over most of the domain, while the right-tail zoom reveals a small but systematic discrepancy.

Quantitative summary measures for representative parameter configurations are reported in Table 1. The discrepancy was also evaluated quantitatively over the plotted grid, using the maximum absolute difference, the root-mean-square difference and the maximum relative difference between the two curves. These measures confirm that the discrepancy is numerically small over most of the domain, but systematic and more pronounced in the right-tail region.

The table confirms the analytical predictions and highlights the most extreme configurations. The largest relative discrepancies occur when the conditional dispersion in the transformed scale is larger and the inverse transformation is more strongly nonlinear. For example, increasing σ_V from 0.5 to 2.0, with $\lambda_Y = 0.5$ and $\rho = 0.8$, substantially increases all distortion measures. Conversely, larger values of ρ reduce the distortion by reducing the conditional dispersion of $V \mid U$. The affine case $\lambda_Y = 1$ provides the benchmark in which the distortion is essentially zero. The scale-free measure $R_{\max}$ is particularly useful for comparing parameter configurations, since it expresses the maximum discrepancy relative to the corresponding original-scale conditional mean.

Table 1. Summary measures of the structural distortion for selected parameter configurations. Here $R_{\max}$ denotes the maximum relative distortion over the numerical grid.

λ_Y	ρ	σ_V	$D_{\max}$	D_{rms}	$R_{\max}$ (Relative)
0.25	0.50	1.00	4.47495×10^{-1}	3.27621×10^{-1}	2.86238×10^{-1}
0.25	0.80	1.00	2.65228×10^{-1}	1.72324×10^{-1}	1.92296×10^{-1}
0.25	0.95	1.00	7.95341×10^{-2}	4.90362×10^{-2}	6.68236×10^{-2}
0.50	0.80	0.50	2.26690×10^{-2}	2.24999×10^{-2}	3.75044×10^{-2}
0.50	0.80	1.00	9.07455×10^{-2}	8.81265×10^{-2}	1.99848×10^{-1}
0.50	0.80	2.00	3.62352×10^{-1}	3.04694×10^{-1}	3.58205×10^{-1}
0.75	0.80	1.00	4.43805×10^{-2}	3.76786×10^{-2}	9.15514×10^{-2}
1.00	0.80	1.00	8.88178×10^{-16}	2.38561×10^{-16}	6.04498×10^{-16}

5. Numerical Sensitivity Analysis

The numerical analysis focuses on the range

$$0 \le \lambda_Y \le 1.$$

This range was chosen in order to include the logarithmic case $\lambda_Y = 0$, genuinely nonlinear positive transformations, and the affine benchmark $\lambda_Y = 1$, for which the structural distortion vanishes. The parameters varied in the numerical experiments are

$$\lambda_Y \in \{0, 0.25, 0.5, 0.75, 1\}, \qquad \rho \in \{0.2, 0.5, 0.8, 0.95\},$$

and

$$\sigma_V \in \{0.5, 1, 1.5, 2\}.$$

Negative values of λ_Y are also relevant in applied Box–Cox modelling, but they lead to an upper-bounded admissible domain

$$\mathcal{D}_{\lambda_Y} = (-\infty, -1/\lambda_Y),$$

and may involve stronger interactions between truncation, boundary behaviour and curvature of the inverse transformation. A systematic study of negative transformation parameters is therefore left for future work. Unless otherwise stated, we set $\lambda_X = 0.3$.

The quantities $M(u)$, $m_T(u)$, and $\tau_T^2(u)$ are evaluated by Monte Carlo integration from the conditional truncated normal distribution.

For each parameter configuration and each value of u on the numerical grid, the Monte Carlo integration was performed using $M = 2 \times 10^5$ simulated draws from the corresponding conditional normal distribution, retaining only draws in the admissible domain $\mathcal{D}_{\lambda_Y}$. The same simulation size was used across all parameter configurations. A fixed random-number seed was used to ensure reproducibility. The computations were carried out in MATLAB R2023a using double-precision arithmetic.

The MATLAB script used for the numerical sensitivity analysis and simulated examples is provided as Supplementary Material as Code S1. The second-order approximation to the distortion is given by

$$\Delta_2(u) = \frac{1}{2}\tau_T^2(u)\, h''_{\lambda_Y}\{m_T(u)\}.$$

Figure 2 shows the effect of λ_Y, with $\rho = 0.8$ and $\sigma_V = 1$. As expected, the distortion decreases as λ_Y approaches the affine case $\lambda_Y = 1$.

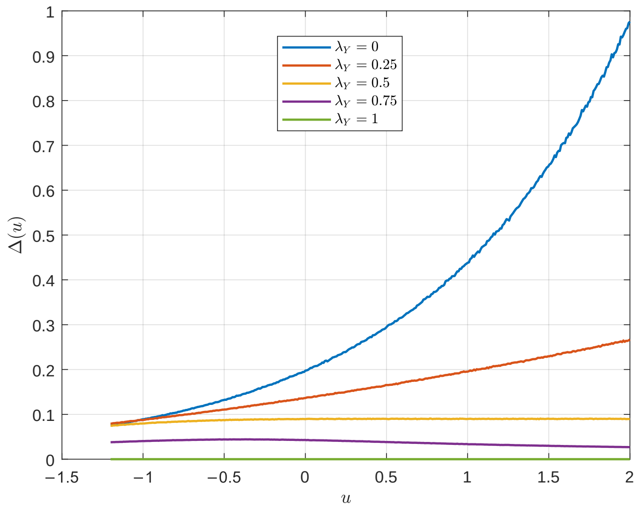

Figure 2. Structural distortion $\Delta(u)$ for different values of $\lambda_Y \in [0, 1]$, with $\rho = 0.8$ and $\sigma_V = 1$.

Figure 3 shows the effect of the correlation parameter. Larger values of ρ reduce the conditional dispersion in the transformed scale and therefore attenuate the curvature-induced distortion.

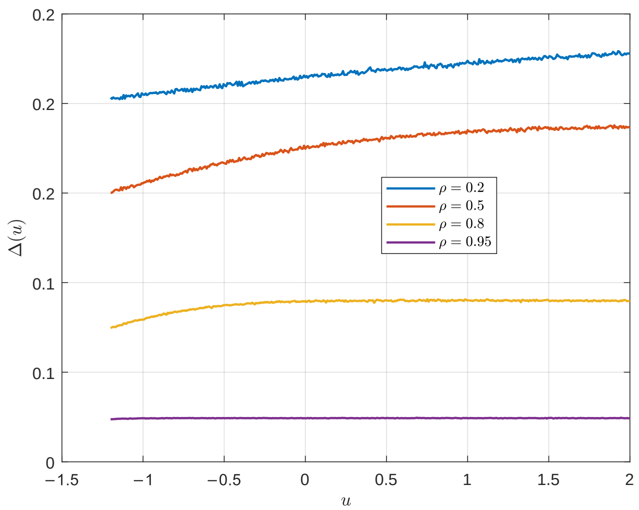

Figure 3. Structural distortion $\Delta(u)$ for different values of ρ, with $\lambda_Y = 0.5$ and $\sigma_V = 1$.

Figure 4 shows the effect of σ_V. Increasing σ_V increases the conditional dispersion of $V \mid U$, thereby amplifying the distortion induced by the nonlinear inverse transformation.

Here $\mathcal{U}$ denotes the grid of admissible u-values used in the numerical evaluation. To summarize the numerical results, Table 1 reports

$$D_{\max} = \max_{u \in \mathcal{U}} |\Delta(u)|, \qquad D_{\mathrm{rms}} = \left\{ \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \Delta(u)^2 \right\}^{1/2},$$

and

$$R_{\max} = \max_{u \in \mathcal{U}} \left| \frac{\Delta(u)}{M(u)} \right|.$$

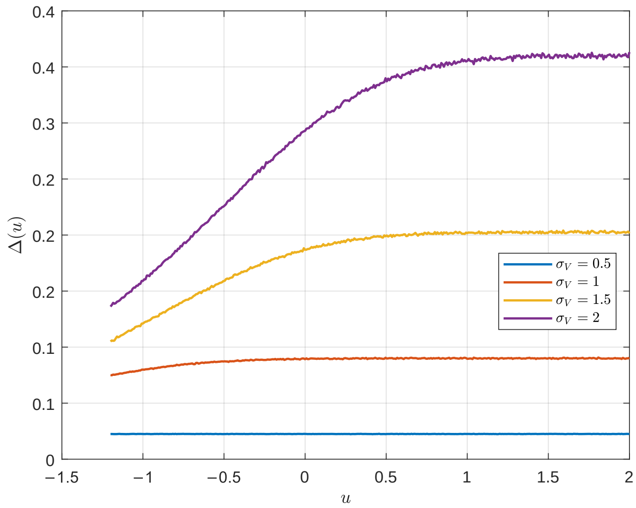

Figure 4. Structural distortion $\Delta(u)$ for different values of σ_V, with $\lambda_Y = 0.5$ and $\rho = 0.8$.

6. A Real-Data Illustration

Let Y denote the variable `charges`. The MATLAB script used for the real-data insurance illustration is provided as Supplementary Material Code S2. Since Y is strictly positive and strongly right-skewed, it provides a natural empirical setting for a Box–Cox transformation. We fitted a Gaussian linear regression model in the transformed response scale. In the main graphical display, BMI was transformed using $\lambda_X = 0.5$, and the model included age, transformed BMI, number of children, smoking status, sex, region, and the interactions age $\times$ smoker and transformed BMI $\times$ smoker as covariates. The response transformation parameter was selected by profile likelihood. The maximum was attained at

$$\widehat{\lambda}_Y = 0.125,$$

indicating a clearly nonlinear transformation of the response variable. For graphical purposes, the fitted curves are displayed as functions of BMI, separately for smokers and non-smokers, with the remaining covariates fixed at reference values. Thus, the figure should be interpreted as a two-dimensional section of the fitted multivariable regression surface, rather than as a simple univariate regression of `charges` on BMI. For each value of BMI in a grid, and separately for smokers and non-smokers, we computed two fitted curves on the original scale. The first is the usual transform–fit–invert curve,

$$\widehat{m}_{\mathrm{TFI}}(x) = g^{-1}_{\widehat{\lambda}_Y}\Big\{\widehat{E}\Big[g_{\widehat{\lambda}_Y}(Y) \mid X = x\Big]\Big\}.$$

The second is the second-order structural approximation to the conditional mean on the original scale,

$$\widehat{m}_{\mathrm{SA}}(x) = g^{-1}_{\widehat{\lambda}_Y}\{\widehat{\mu}(x)\} + \frac{1}{2}\Big(g^{-1}_{\widehat{\lambda}_Y}\Big)''\{\widehat{\mu}(x)\}\widehat{\sigma}^2,$$

where $\widehat{\mu}(x)$ is the fitted conditional mean in the transformed scale and $\widehat{\sigma}^2$ is the residual variance in that scale. For the Box–Cox inverse transformation,

$$g^{-1}_{\lambda}(z) = (1 + \lambda z)^{1/\lambda},$$

the second derivative is

$$\Big(g^{-1}_{\lambda}\Big)''(z) = (1 - \lambda)(1 + \lambda z)^{1/\lambda - 2}.$$

Figure 5 compares the usual transform–fit–invert curves with the corresponding structural second-order approximations. The two curves are close, but they are not identical. The discrepancy is systematic and has the same origin as in the theoretical development: the nonlinear inverse Box–Cox transformation does not commute with conditional expectation.

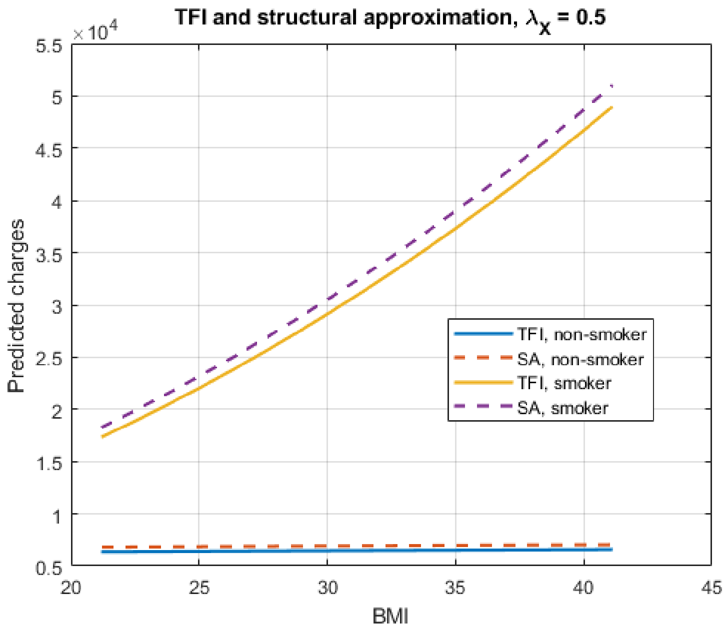

Figure 5. Medical insurance charges data. Comparison between the usual transform–fit–invert predictor and the second-order structural approximation to the conditional mean on the original scale, as a function of BMI and smoking status, for $\lambda_X = 0.5$. The displayed curves are two-dimensional sections of a multivariable fitted model. The example is intended only as an empirical illustration of the structural discrepancy, not as an optimal predictive model for medical charges.

The relative discrepancy

$$100\,\frac{\widehat{m}_{\mathrm{SA}}(x) - \widehat{m}_{\mathrm{TFI}}(x)}{\widehat{m}_{\mathrm{TFI}}(x)}$$

is shown in Figure 6. Over the displayed BMI range, the structural approximation differs from the usual transform–fit–invert predictor by approximately 4% to 7%. For non-smokers, the discrepancy is nearly constant, around 6.5%, whereas for smokers it ranges from slightly above 5% at lower BMI values to approximately 4% at higher BMI values.

To examine whether the transformation of the explanatory variable affects the magnitude of the discrepancy, we performed a sensitivity analysis with respect to λ_X. The response transformation was kept fixed at the profile-likelihood estimate $\widehat{\lambda}_Y = 0.125$, while the BMI transformation parameter was varied over the grid

$$\lambda_X \in \{0, 0.125, 0.25, 0.5, 0.75, 1\}.$$

For each value of λ_X, the regression model was refitted in the transformed scale, and the relative discrepancy between the transform–fit–invert predictor and the second-order structural approximation was computed over the displayed BMI range and for both smoking groups. The results are reported in Table 2.

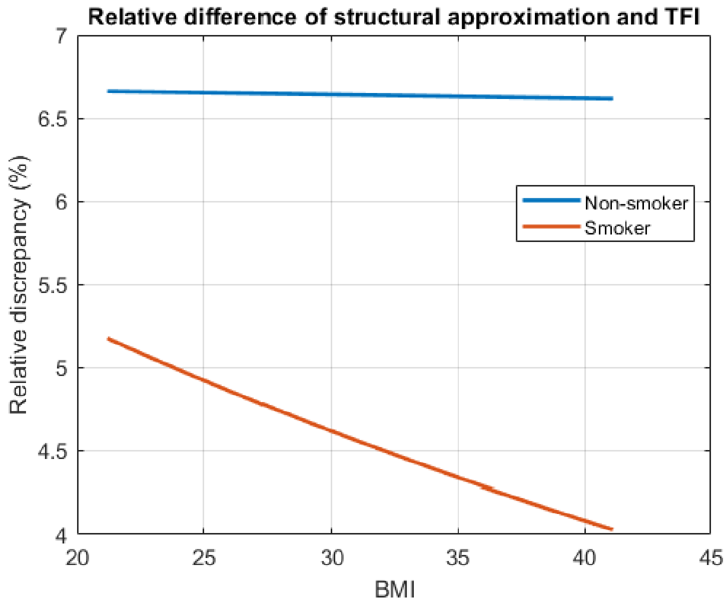

Figure 6. Relative difference between the second-order structural approximation and the usual transform–fit–invert predictor in the medical insurance charges data.

Table 2. Sensitivity of the relative discrepancy to different values of λ_X, with $\lambda_Y = 0.125$.

λ_X	Minimum	Mean	Maximum	Standard Deviation
0	4.1909	5.8075	6.9434	1.1302
0.125	4.1849	5.8087	6.9427	1.1301
0.250	4.1799	5.8107	6.9430	1.1301
0.500	4.1732	5.8175	6.9471	1.1304
0.750	4.1706	5.8280	6.9554	1.1313
1.000	4.1724	5.8419	6.9680	1.1328

In this empirical example, the effect of λ_X on the magnitude of the discrepancy is limited. The mean relative discrepancy varies only from approximately 5.81% to 5.84%, while the maximum discrepancy remains close to 7% for all values of λ_X. Thus, although the transformation of the explanatory variable can slightly modify the numerical magnitude of the discrepancy, it does not remove its structural origin. The discrepancy remains a consequence of applying a nonlinear inverse transformation to a conditional summary computed in the transformed response scale. This empirical illustration supports the main message of the paper. Even in a simple real-data regression setting, the curve obtained by fitting a linear Gaussian model in the transformed scale and then applying the inverse Box–Cox transformation should not automatically be interpreted as the conditional mean curve on the original scale. The observed discrepancy is not a numerical artefact of the theoretical examples nor a consequence of a particular choice of λ_X in this illustration; it is a structural consequence of applying a nonlinear inverse transformation to a conditional summary computed in another scale.

7. Practical Implications and Conclusions

The main practical implication is that an inverse-transformed fitted value should not automatically be interpreted as a conditional mean on the original scale. In the admissible-domain formulation, the inverse-transformed fitted curve is

$$h_{\lambda_Y}\{m_T(u)\},$$

whereas the conditional mean of the original response is

$$\mathbb{E}[h_{\lambda_Y}(V) \mid U = u,\ V \in \mathcal{D}_{\lambda_Y}].$$

These quantities coincide only under special conditions, such as an affine inverse transformation or a degenerate conditional distribution. In general, they answer different statistical questions: the former is the inverse image of a transformed-scale conditional mean, whereas the latter is the mean predictor under squared-error loss on the original scale.

The second-order expansion provides a diagnostic for this discrepancy:

$$C_T(u) = \frac{1}{2} h''_{\lambda_Y}\{m_T(u)\}\tau_T^2(u).$$

Thus transformation-based regression should distinguish between the transformed-scale conditional mean, the inverse-transformed fitted curve, and the original-scale conditional mean. Conflating these objects may lead to incorrect interpretation of regression curves, fitted values, conditional effects, and uncertainty summaries on the original scale.

The analysis also clarifies the relation with classical retransformation bias. Retransformation bias is usually concerned with correcting predictions after returning from a transformed scale to the original scale. The perspective developed here is structural: nonlinear inverse transformations do not, in general, preserve the conditional mean operator. This discrepancy is not a finite-sample effect, an estimation error, or a numerical artefact.

Although the paper focuses on conditional means, the same invariance question is relevant for other regression summaries. The answer depends on the statistical functional being considered. Conditional quantiles behave differently from means: under strictly monotone transformations, quantiles are equivariant, so that the inverse transformation of a transformed-scale conditional quantile corresponds to the conditional quantile on the original scale, provided that the conditioning structure is treated consistently. In contrast, conditional variances, prediction intervals and other uncertainty summaries are not generally preserved in a simple way by nonlinear inverse transformations, because such transformations change scale, curvature and dispersion asymmetrically. Thus, the present analysis should be viewed as the conditional-mean case of a broader invariance problem: which statistical summaries are preserved, and which are structurally altered, by transform–fit–inverse modelling procedures. A systematic treatment of quantiles, variances, prediction intervals and other regression summaries lies beyond the scope of the present paper, but represents a natural direction for future work.

The results do not argue against nonlinear transformations. Transformations such as Box–Cox may remain useful for improving normality, stabilizing variance, or simplifying dependence structures. However, once fitted relationships are mapped back to the original scale, the target of interpretation changes unless the non-commutativity with conditional expectation is explicitly accounted for.

From a practical point of view, the structural approximation is most relevant when the target of inference is the conditional mean of the response on the original scale. This includes situations in which inverse-transformed fitted values are used as mean predictions, regression curves are interpreted on the original scale, or covariate effects are discussed after back-transformation. In such cases, the transform–fit–invert curve and the original-scale conditional mean represent different statistical objects.

The magnitude of the discrepancy is expected to be larger when the inverse transformation is strongly nonlinear over the relevant range and when the conditional dispersion in the transformed scale is not negligible. In contrast, if the inverse transformation is nearly affine over the region of interest, or if the transformed-scale conditional variance is very small, the discrepancy may be numerically minor. Practitioners should therefore distinguish between using a transformation as a modelling device in the transformed scale and interpreting inverse-transformed fitted values as mean predictions on the original scale.

Future work may proceed in several directions. One direction is to develop higher-order correction formulas and assess their finite-sample behaviour. Another is to study analogous structural distortions for other transformation families, including transformations designed for zero or negative data. More broadly, the non-commutativity framework may be extended beyond Box–Cox models to other nonlinear transformations, filtering operations, and dependence structures.

Overall, the paper shows that inverse Box–Cox distortion is not merely a technical bias-correction issue. It reflects a structural limitation of nonlinear transformation-based modelling: transformations may simplify the data in one scale, but they do not generally preserve conditional mean structure when mapped back to the original scale.

Supplementary Materials: The following supporting information is available online: https://www.mdpi.com/article/10.3390/axioms15070519/s1, Code S1: MATLAB script for the numerical sensitivity analysis and simulated examples, including all required local functions; Code S2: MATLAB script for the real-data insurance illustration, including all required local functions; Text S1: README file describing the supplementary materials.

Funding: This research received no external funding.

Data Availability Statement: The medical insurance charges dataset used in this study is publicly available online from the Kaggle repository, *Medical Cost Personal Datasets*, at https://www.kaggle.com/datasets/mirichoi0218/insurance (accessed on 5 July 2026). The numerical illustrations were based on simulated data generated using MATLAB R2023a scripts. The MATLAB scripts used to generate the simulated data, numerical sensitivity analysis, real-data illustration, and figures are provided as Supplementary Materials accompanying this manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Box, G.E.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211–243. [CrossRef]
2. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 1st ed.; Wiley: Hoboken, NJ, USA, 1966.
3. Carroll, R.J.; Ruppert, D. *Transformation and Weighting in Regression*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1988.
4. Sakia, R.M. The Box–Cox Transformation Technique: A Review. *Statistician* **1992**, *41*, 169–178. [CrossRef]
5. Goto, M.; Inoue, T. Some Properties of the Power-Normal Distribution. *J. Biom.* **1980**, *1*, 28–54. [CrossRef]
6. Freeman, J.; Modarres, R. Inverse Box-Cox: The Power-Normal Distribution. *Stat. Probab. Lett.* **2006**, *76*, 764–772. [CrossRef]
7. Goto, M.; Inoue, T.; Tsuchya, Y. On the Estimation of Parameters in the Power-Normal Distribution. *Bull. Inf. Cybern.* **1984**, *21*, 41–53. [CrossRef]
8. Goto, M.; Hamasaki, T. The Bivariate Power-Normal Distribution. *Bull. Inform. Cybernet.* **2002**, *34*, 29–49. [CrossRef]
9. Blum, L.; Elgendi, M.; Menon, C. Impact of Box-Cox Transformation on Machine-Learning Algorithms. *Front. Artif. Intell.* **2022**, *5*, 877569. [CrossRef] [PubMed]
10. Alshamrani, A. Machine Learning Techniques Improving the Box–Cox Transformation in Breast Cancer Prediction. *Electronics* **2025**, *14*, 3173. [CrossRef]
11. Duan, N. Smearing Estimate: A Nonparametric Retransformation Method. *J. Am. Stat. Assoc.* **1983**, *78*, 605–610. [CrossRef]
12. Miller, D.M. Reducing Transformation Bias in Curve Fitting. *Am. Stat.* **1984**, *38*, 124–126. [CrossRef]
13. Taylor, J.M.G. The Retransformed Mean after a Fitted Power Transformation. *J. Am. Stat. Assoc.* **1986**, *81*, 114–118. [CrossRef]
14. Sakia, R.M. Retransformation Bias: A Look at the Box–Cox Transformation to Linear Models. *Commun. Stat. Simul. Comput.* **1990**, *19*, 189–203.
15. Manning, W.G.; Mullahy, J. Estimating Log Models: To Transform or Not to Transform? *J. Health Econ.* **2001**, *20*, 461–494. [CrossRef]
16. Hoyle, M.H. Transformations: An Introduction and a Bibliography. *Int. Stat. Rev.* **1973**, *41*, 203–223. [CrossRef]

17. Yeo, I.-K.; Johnson, R.A. A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika* **2000**, *87*, 954–959. [CrossRef]
18. Velilla, S. A Note on the Multivariate Box–Cox Transformation to Normality. *Stat. Probab. Lett.* **1993**, *17*, 259–263. [CrossRef]