

Beyond p-values: A comparative framework for estimation-based inference using the a priori procedure and gain–probability analysis

Tonghui Wang

Department of Mathematics, New Mexico State University, USA

S.T. Boris Choy

Discipline of Business Analytics, The University of Sydney, Australia

David Trafimow

Department of Psychology, New Mexico State University, USA

Xiangfei Chen

Department of Mathematics, Bridgewater State University, USA

and Ziyuan Wang

Department of Mathematics, University of Wisconsin-Oshkosh, USA

Abstract

Purpose - This paper compares three inferential paradigms, Null Hypothesis Significance Testing (NHST), the A Priori Procedure (APP) and Gain–Probability (G–P) analysis, and advocates a conceptual shift away from binary hypothesis testing toward inference frameworks grounded in estimation and probability.

Design/methodology/approach - NHST, which relies on p-values and dichotomous decision rules, is contrasted with APP, which emphasizes pre-data estimation reliability, and with G–P analysis, which quantifies the probability that one outcome exceeds another and the extent to which it does. APP is derived for parameter estimation in linear regression models under skew-normal error structures, including results on the distribution of regression estimators. The G–P framework is examined for gamma-distributed variables in both independent and dependent settings. Both are supported by simulation studies and real-world applications.

Findings - Beyond their conceptual advantages, the APP and G–P paradigms are highly applicable in practice, as shown by the simulation studies and empirical illustrations. The shift they represent aligns with broader methodological reforms that prioritize estimation, transparency and interpretability.

Originality/value - The paper extends APP to linear regression with skew-normal errors and G–P analysis to gamma-distributed variables under independence and dependence. Together, these frameworks offer a rigorous and interpretable alternative to conventional significance testing by prioritizing estimation accuracy, probabilistic interpretation and practical relevance.

Keywords Null Hypothesis Significance Testing (NHST), A Priority Procedure (APP), Gain Probability (G-P) Analysis, Estimation Based Inference, Skew Normal distribution, Gamma Distribution, Statistical Methodology.

Paper type Research paper

1. Introduction

Null Hypothesis Significance Testing (NHST) has long served as the dominant inferential framework across disciplines. Despite its historical prominence, NHST has attracted growing criticism due to its reliance on p-values and post hoc binary decision rules, which often encourage dichotomous thinking and undermine interpretability. The ongoing reproducibility crisis in psychology, medicine, and the social sciences has further intensified calls for methodological

JEL Classification —C12, C13, C18, C46

© Tonghui Wang, S.T. Boris Choy, David Trafimow, Xiangfei Chen and Ziyuan Wang. Publish in Asian Journal of Applied Financial Econometrics.

reform, advocating for inference approaches that emphasize estimation, uncertainty quantification, and transparency (Wasserstein and Lazar (2016), Trafimow and Marks (2015)). The editorial in the 2019 issue of *The American Statistician* actually calls for NHST to be abandoned (Wasserstein et al. (2019)).

In response to these challenges, alternative paradigms have been developed that move beyond the limitations of NHST. Two such frameworks are the A Priori Procedure (APP) and Gain–Probability (G–P) analysis. The APP, initially advanced to address sample size determination in estimation contexts, focuses on ensuring estimation reliability prior to data collection. It circumvents the issues associated with post hoc inference by explicitly targeting accuracy and precision in parameter estimation. In parallel, G–P analysis offers a probabilistic framework for assessing the likelihood that one outcome exceeds another, and by varying magnitudes, thereby replacing dichotomous significance judgments with interpretable comparative probabilities. These developments, as elaborated in Tong et al. (2022b) and Trafimow et al. (2019), Trafimow et al. (2023), Trafimow et al. (2024) represent a broader shift toward estimation-based and probabilistic inference that is increasingly endorsed by reform-oriented statisticians and methodologists.

This paper undertakes a comparative and integrative study of NHST, APP, and G–P analysis, situating each within the broader context of inferential reasoning and reproducibility reform. In particular, we explore the theoretical foundations and practical merits of APP and G–P analysis as coherent alternatives without of using traditional significance testing. From a methodological standpoint, we detail the derivation of APP procedures for linear regression models under skew-normal error distributions, including analysis of estimator distributions, simulation-based evaluations, and real-data applications. From a probabilistic perspective, we examine the G–P framework in the context of gamma-distributed outcomes, addressing both independent and dependent scenarios.

By integrating these inferential strategies, this work aims to provide a unified perspective that emphasizes estimation accuracy, probabilistic interpretation, and practical effect understanding. Our results contribute to the growing body of literature seeking to reform and enrich the foundations of statistical inference in applied research.

2. The Classical NHST Framework

The NHST framework has been the dominant inferential paradigm in scientific research for much of the twentieth century. It combines elements of Fisher’s significance testing and the Neyman–Pearson decision-theoretic approach into a hybrid procedure that is now widely taught and practiced across disciplines (Gigerenzer (2004); Hubbard and Bayarri (2003)).

Under NHST, researchers begin with a null hypothesis—typically representing a statement of no effect, no difference, or no association—and evaluate it against an alternative hypothesis using a test statistic derived from the observed data. A central quantity in this framework is the p-value, defined as.

$$p = \Pr(\text{data or more extreme} \mid H_0). \quad (2.1)$$

This p-value represents the probability of obtaining results as extreme as those observed, assuming that H_0 is true. If the computed p-value falls below a predetermined significance level α (commonly set at 0.05), the result is deemed “statistically significant,” and H_0 is rejected in favor of H_1 .

Despite its formal simplicity, NHST has been the subject of sustained criticism. A primary issue is the widespread misinterpretation of p-values. They are often mistakenly viewed as the probability that the null hypothesis is true, or as a measure of the effect’s importance or replicability. In reality, the p-value is conditional on the assumption that H_0 is true and does not provide a direct probability statement about itself (Goodman (2008)).

Moreover, p-values are highly sensitive to sample size. With large samples, even negligible effects may yield small p-values, leading to the rejection of H_0 despite practical irrelevance. Conversely, small samples may fail to detect substantively meaningful effects due to insufficient statistical power (Cohen (1994)). This dependence on sample size can lead to inconsistent or misleading conclusions, especially in exploratory or observational studies.

Another fundamental limitation is that NHST provides no direct information about the magnitude, direction, or practical significance of an effect. The binary outcome of “significant” vs. “non-significant” can foster a false sense of certainty and promotes dichotomous thinking, which can obscure the underlying uncertainty and complexity of empirical phenomena (Amrhein et al. (2019), Trafimow et al. (2019)).

In light of these limitations, many scholars have called for the replacement or supplementation of NHST with approaches that emphasize estimation, effect sizes, confidence intervals, and probability-based reasoning. These reforms aim to foster more transparent, interpretable, and reproducible scientific practices (Wasserstein and Lazar (2016); Cumming (2014)). In this paper, our proposed estimation based procedures can be the replacements of NHST, which will be detailed in next two sections.

3. The A Priori Procedure (APP)

The APP redefines inference by emphasizing estimation precision rather than significance. Instead of post hoc testing, APP determines the sample size needed for estimating and testing the population parameters with adequate precision using good sample statistics. Desired properties of good estimators include unbiasedness, minimum variance or mean square error (MSE), consistency, and robustness. The APP is guided by two fundamental questions:

- Precision (f): How close do we want the sample statistics to be to their corresponding population parameters?
- Confidence level (c): With what probability do we want to have in order to meet the precision specification?

For example, a researcher might wish to have a 95% probability ($c=0.95$) of obtaining a sample mean within a tenth precision ($f=0.1$) of a standard deviation of the population mean. Specifically, APP determines the minimum sample size n required to ensure an estimator $\hat{\theta}$ of the population parameter θ achieves a desired level of accuracy:

$$\Pr\left(\left|\hat{\theta} - \theta\right| < \delta\right) \geq 1 - \alpha, \quad (3.1)$$

where δ is the researcher-specified tolerance and $1 - \alpha$ is the assurance probability. APP thus guarantees reliability prior to data collection, aligning statistical inference with the goals of replicability and precision.

In the following, we will use the APP procedure for estimating parameters in regression model with skew normal error structures.

3.1 Linear Regression model with skew normal errors

The multiple linear regression model for n observations is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n,$$

which is equivalent to the matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \quad (3.2)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$ is the vector of dependent variable,

$$\mathbf{X} \equiv (\mathbf{j}_n, \mathbf{X}_1) = (\mathbf{j}_n, \mathbf{x}_1, \dots, \mathbf{x}_p)$$

is $n \times (p + 1)$ design matrix with $\mathbf{j}_n = (1, \dots, 1)'$ in $\mathfrak{R}^n$.

Proposition 1. (Tong et al. (2022a)) Consider the linear regression model given in (3.2) and assume that $\mathbf{E} \sim SN_n(0, \omega^2 \mathbf{I}_n, \boldsymbol{\alpha})$. Then the least squares estimators of $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \dots, \beta_p)$, and ω^2 are, respectively, given by

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c' \mathbf{y}, \quad s^2 \equiv \hat{\omega}^2 = \frac{\mathbf{y}'(\mathbf{I}_n - \mathbf{H})\mathbf{y}}{n - p - 1}, \quad (3.3)$$

where $\mathbf{X}_c = (\mathbf{I}_n - \bar{\mathbf{J}}_n) \mathbf{X}_1$ with $\bar{\mathbf{J}}_n = \frac{1}{n} \mathbf{j}_n \mathbf{j}_n'$, and $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$. Furthermore, these estimators have the following properties.

(i) $\hat{\boldsymbol{\beta}}_1 \sim SN_p(\boldsymbol{\beta}_1, \omega^2 (\mathbf{X}_c' \mathbf{X}_c)^{-1}, \boldsymbol{\alpha}_1)$, where $\boldsymbol{\alpha}_1 = (\mathbf{X}_c' \mathbf{X}_c)^{-1/2} \mathbf{X}_c' \boldsymbol{\alpha} / \sqrt{1 + \boldsymbol{\alpha}'(\mathbf{I}_n - \mathbf{H})\boldsymbol{\alpha}}$.

(ii) $(n - p - 1)s^2 / \omega^2 \sim \chi_{n-p-1}^2$, the chi-squared distribution with $n - p - 1$ degrees of freedom.

(iii) $\hat{\boldsymbol{\beta}}_1$ and s^2 are independent.

3.2 The APP Procedure for estimating regression coefficients

Theorem 1. Assume that $y \sim SN_n(X\beta, \omega^2 I_n, \alpha)$, where $\mathbf{X}$ is $n \times (p+1)$ of rank $p+1 < n$ and $\beta = (\beta_0, \beta_1, \dots, \beta_p)' = (\beta_0, \beta_1)'$. Let β_{10} be the true regression coefficients. Then for specified precision f and confidence level c , the necessary sample size needed for estimating regression coefficients $\hat{\beta}_1$ (or regression weights) can be obtained by solving the following equation

$$\int_0^{(n-1)f^2} f_U(u) du = c, \quad (3.4)$$

where

$$U \equiv \frac{S_{xx}^{1/2} (\hat{\beta}_1 - \beta_{10})^2}{ps^2 / (n-1)} \sim SF_{p, n-p-1}(\lambda, \delta_1, \delta_2), \quad (3.5)$$

the skew F-distribution with numerator degrees of freedom p and denominator degrees of freedom $n-p-1$, the noncentral parameter λ , and the skewness parameters δ_1 and δ_2 , and $f_U(u)$ is the density function of U . Here

$$\lambda = \frac{(\beta_1 - \beta_{10})'(X_c' X_c)(\beta_1 - \beta_{10})}{\omega^2}, \quad \delta_1 = \frac{(\beta_1 - \beta_{10})' X_c' \alpha}{\omega \sqrt{1 + \alpha'((I_n - H_c))\alpha}},$$

$$\delta_2 = \frac{\alpha' H_c \alpha}{1 + \alpha' (I_n - H_c) \alpha}$$

and $S_{xx} = (X_c' X_c)$.

Proof. From (ii) and (iii) of Proposition 1, we have $(n-p-1)s^2 / \omega^2 \sim \chi_{(n-p-1)}^2$ and $\hat{\beta}_1, s^2$ are independent. By the definition of the skew F-distribution given in Ma et al. (2022), it follows that

$$U \equiv \frac{\|S_{xx}^{1/2} (\hat{\beta}_1 - \beta_{10})\|^2}{ps^2 / (n-1)} \sim SF_{p, n-p-1}(\lambda, \delta_1, \delta_2) \quad (3.6)$$

the skew F-distribution with numerator degrees of freedom p and denominator degrees of freedom $n-p-1$, the noncentral parameter λ , and the skewness parameters δ_1 and δ_2 . The density function of U is given by

$$f_U(u; \lambda, \delta_1, \delta_2) = \frac{p e^{-\lambda/2}}{(n-p-1) \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{p-1}{2}\right) \Gamma\left(\frac{n-p-1}{2}\right) 2^{\frac{n-1}{2}-1}} h_1(u; \lambda, \delta_1, \delta_2), \quad u > 0,$$

where

$$h_1(u; \lambda, \delta_1, \delta_2) = \int_0^\infty \int_{-\sqrt{\frac{put}{n-p-1}}}^{\sqrt{\frac{put}{n-p-1}}} \exp\left\{-\frac{1}{2}\left[t\left(1 + \frac{p}{n-p-1}u\right) + \lambda^{1/2}v\right]\right\}$$

$$\times \left(\frac{p}{n-p-1}ut - v^2\right)^{\frac{p-3}{2}} t^{(n-p-1)/2} \Phi\{\alpha_0(v - \lambda^{1/2})\} dv dt, \quad \alpha_0 = \frac{\lambda^{-1/2} \delta_1}{\left(1 + \delta_2 - \frac{\delta_1^2}{2}\right)^{1/2}},$$

and $\Gamma(\cdot)$ is the gamma function.

Assume that $E \sim SN_n(0, \omega^2 I_n, \alpha)$, then for specified precision f and confidence level c , the necessary sample size needed for estimating regression coefficients $\hat{\beta}_1$ can be obtained by solving the following equation

$$\int_0^{(n-1)f^2} f_U(u) du = c,$$

where $f_U(u)$ is given above.

3.3 Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the APP proposed above. The necessary sample sizes (n) obtained by using Theorem 1 are provided in Table 1 and Table 2 for various combinations of p , λ , δ_1 and δ_2 when $f = 0.15, 0.2, 0.25$, and $c = 0.9, 0.95$.

Table 1. The necessary sample sizes for $\lambda=2$, $\delta_1=0.5$, $\delta_2=1$

f	0.15		0.2		0.25	
c	0.9	0.95	0.9	0.95	0.9	0.95
p = 2, n	113	139	66	81	44	54
cr	0.9007	0.9503	0.9015	0.9487	0.9046	0.9526
p = 4, n	77	91	47	55	33	39
cr	0.9011	0.9510	0.8988	0.9517	0.9033	0.9554
p = 6, n	66	76	42	48	30	35
cr	0.9003	0.9509	0.9051	0.9489	0.9061	0.9548
p = 12, n	59	65	41	45	32	35
cr	0.8994	0.9512	0.9009	0.9512	0.9006	0.9532

Source(s): Table by the authors

Table 2. The necessary sample sizes and coverage rates for $p=2,4,6,12$, $\lambda=2$, $\delta_1=0.5$, $\delta_2=1$

f	0.15		0.2		0.25	
c	0.9	0.95	0.9	0.95	0.9	0.95
p=2, n	113	139	66	81	44	54
cr	0.9007	0.9503	0.9015	0.9487	0.9046	0.9526
p=4, n	77	91	47	55	33	39
cr	0.9011	0.9510	0.8988	0.9517	0.9033	0.9554
p=6, n	66	76	42	48	30	35
cr	0.9003	0.9509	0.9051	0.9489	0.9061	0.9548
p=12, n	59	65	41	45	32	35
cr	0.8994	0.9512	0.9009	0.9512	0.9006	0.9532

Source(s): Table by the authors

For calculating the necessary sample sizes needed to estimate the regression coefficients β_1 , we provide a freely available online calculator at: <https://appcohensd.shinyapps.io/appregressionskew/>.

3.4 Real Data Example

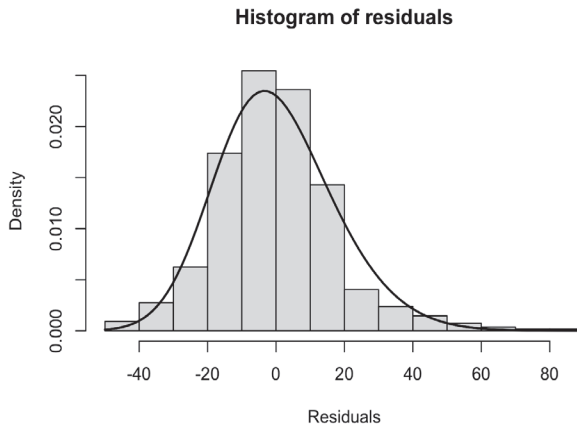
The data used in this example come from the HousePrices data set in the R package AER, HousePrices, which contains information on housing prices and housing characteristics. The response variable is the sale price of a house, while the explanatory variables considered are lot size, number of bedrooms, number of bathrooms, number of stories, and number of garage places. To improve numerical stability and interpretability, the housing price was measured in thousands of dollars and the lot size was measured in thousands of square feet. We construct a regression model to predict housing price (Y) on the basis of lot size (x_1), bedrooms (x_2), bathrooms (x_3), stories (x_4), and garage places (x_5). Thus, $p = 5$ in this example.

If we use the entire data set as a population ($N=546$), the least squares estimator of β_1 and standardized coefficients β_{1*} are given by, respectively,

$$\beta_{10} = (4.73, 2.39, 16.15, 7.88, 5.37)' \text{ and } \beta_{1*} = (0.38, 0.07, 0.30, 0.25, 0.17)'.$$

To obtain the estimations of λ , δ_1 , and δ_2 , we first fitted a skew-normal distribution to the residuals from the regression model based on the whole data set in Figure 1. The fitted parameters of skew normal distribution were $\xi = -16.42$, $\omega = 24.12$, $\alpha = 1.71$.

Then, using the pilot sample of size $n_0 = 100$ and treating β_{10} as the empirical population



Source(s): Figure by the authors

Figure 1. Histogram and density curve of the residuals from the housing price regression model by the fitted skew-normal distribution using the estimated location vector $\xi = -16.421_n$, scale matrix $\omega^2 = (24.12)^2 I_n$, and shape vector $\alpha = 1.711_n$.

value, the estimates of λ , δ_1 , and δ_2 were obtained from the plug-in formulas in Theorem 1. Specifically, λ was estimated from the quadratic form involving $\hat{\beta}_{1,plug} - \hat{\beta}_{10}$, the centered design matrix, and the whole-data estimate of ω^2 . This gives that $\hat{\lambda} = 4.88$, $\hat{\delta}_1 = 0$, $\hat{\delta}_2 = 0$.

Using the online calculator provided above, with $p = 5$, precision $f = 0.1$, confidence level $c = 0.95$, and the above estimates of λ , δ_1 , and δ_2 , we obtain the necessary sample size needed for estimating β_1 as $n = 219$. So we randomly choose a sample of size $n = 219$ from the original data set. After calculation, the least squares estimate of β_1 and $\hat{\beta}_1^*$ are, respectively,

$$\hat{\beta}_1 = (5.39, 3.42, 14.30, 6.85, 5.96)' \text{ and } \hat{\beta}_1^* = (0.41, 0.10, 0.26, 0.23, 0.19)'.$$

For comparison, the difference between β_{10} and $\hat{\beta}_1$ is

$$\hat{\beta}_1 - \beta_{10} = (0.65, 1.03, -1.84, -1.03, 0.58)'.$$

Since the regression coefficients are measured on different scales, we also compare the standardized regression weights. The difference between β_{10}^* and $\hat{\beta}_1^*$ is

$$\hat{\beta}_1^* - \beta_{10}^* = (0.02, 0.03, -0.04, -0.03, 0.02)'.$$

The largest absolute difference among the standardized regression weights is approximately 0.04, which suggests that the proposed APP procedure with the required sample size $n = 219$ provides estimates close to the empirical whole-data estimates in this application.

4. Gain-Probability (G-P) Analysis

The Gain-Probability (G-P) framework extends estimation-based reasoning to comparative contexts. For two populations X and Y , the gain probability is defined as

$$G = \Pr(X > Y), \quad (4.1)$$

representing the probability that a randomly selected observation from X exceeds one from Y . Unlike p-values, which quantify extremeness under a null, G directly expresses the probabilistic meaning of a difference. For example, $G = 0.70$ implies a 70% chance that an observation from one condition outperforms another, providing both interpretability and effect magnitude. Although G-P analyses have been developed for a range of distributions, including normal, skew-normal, log-normal, log skew-normal, delta log skew-normal, and gamma distributions, we focus here on the normal case, consistent with the assumption of bivariate normality. For example, imagine that large and randomly selected samples of people from Culture X and Culture Y complete a test of happiness, with the following means and standard deviations.

- Culture X : Mean = 60, Standard deviation = 10, normal distribution.
- Culture Y : Mean = 55, Standard deviation = 10, normal distribution.

The corresponding Cohen's d is 0.5, typically interpreted as a medium effect size. But the researcher might wish to estimate the probability that a randomly selected person from Culture X would be happier than a randomly selected person from Culture Y . Or, a researcher might desire even more precise information, such as the probability that a randomly selected person from Culture X will be happier than a randomly selected person from Culture Y by between 10 and 20 happiness points or some other interval. G-P analyses can provide all this information.

In the following, we will illustrate the method of G-P Analysis for family of gamma distribution in detail (Wang et al. 2025).

4.1 A Gain-Probability Framework for Gamma-Distributed Data in Risk-Based Decision Making
Gain-probability (G-P) analyses and diagrams provide estimates of the probability of experiencing varying degrees of gain or loss for one group relative to another. Such comparisons may involve experimental versus control groups, females versus males, collectivist versus individualist cultures, and many other contexts. Trafimow et al. (2024) illustrated this approach using competing treatments for high blood pressure, where conventional statistical reporting typically includes means, standard deviations, and p-values. Unfortunately, such information alone is insufficient for clinicians to make informed decisions about whether to prescribe a new treatment. In medical settings, for example, clinicians need to know the probability that a patient will be better or worse off under a new treatment compared to the standard treatment, and by how much. G-P analyses and diagrams provide exactly this type of probabilistic and magnitude-based information, thereby offering a more intuitive and decision-relevant framework.

The family gamma distributions, characterized by scale and shape parameters, offers remarkable flexibility and encompasses several important distributions, including the exponential, Erlang, and chi-squared distributions, as special cases. In this subsection we develop the G-P framework for the gamma distributions, thereby extending its applicability to these related distributions as well.

4.2 Family of Gamma Distributions

The gamma distribution, denoted $G(\alpha, \beta)$, is a continuous two-parameter family with probability density function given by

$$f_G(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), x > 0, \quad (4.2)$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter, and $\Gamma(\cdot)$ denotes the gamma function.

A wide variety of bivariate gamma distributions have been developed for diverse modeling purposes. Chapter 48 of Kotz et al. (2019) provides a comprehensive overview.

For the bivariate gamma Distribution, we adopt the formulation proposed by Nadarajah and Gupta (2006), which possesses analytically tractable marginal distributions and is particularly well-suited for modeling dependent, positively skewed data.

Let V_1 , V_2 and W be independent, absolutely continuous, non-negative random variables. Suppose that W follows a beta distribution with shape parameters α_1 and α_2 and that

$$V_1 \sim G(\alpha_1 + \alpha_2, \beta_1), \quad V_2 \sim G(\alpha_1 + \alpha_2, \beta_2).$$

The resulting bivariate gamma distribution is denoted $BG(\alpha_1, \alpha_2, \beta_1, \beta_2)$. Define

$$X_1 = V_1 W, \quad X_2 = V_2 W.$$

Then the joint distribution of $\mathbf{X} = (X_1, X_2)^\top$ has density

$$f_{BG}(x_1, x_2) = \frac{\left(\frac{x_1}{\beta_1} + \frac{x_2}{\beta_2}\right)^{-\alpha_1 + \frac{1}{2}\alpha_2} (x_1 x_2)^{\alpha_1 + \alpha_2 - 1}}{\beta_1^{\alpha_1 + \alpha_2} \beta_2^{\alpha_1 + \alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_1 + \alpha_2)} \times \exp\left\{-\frac{1}{2}\left(\frac{x_1}{\beta_1} + \frac{x_2}{\beta_2}\right)\right\} W_{\alpha_1 + \frac{1}{2}\alpha_2, \alpha_2} \left(\frac{x_1}{\beta_1} + \frac{x_2}{\beta_2}\right), \quad (4.3)$$

for $x_1 > 0$ and $x_2 > 0$. Here, $W_{\lambda, \mu}(y)$ denotes the Whittaker function, which can be defined by

$$W_{\lambda, \mu}(y) = \frac{y^{\mu+\frac{1}{2}} \exp(-y/2)}{\Gamma\left(\mu - \lambda + \frac{1}{2}\right)} \int_0^\infty t^{\mu-\lambda-\frac{1}{2}} (1+t)^{\mu+\lambda-\frac{1}{2}} \exp(-yt) dt. \quad (4.4)$$

An important property of this distribution is that its marginals are gamma:

$$X_1 \sim G(\alpha_1, \beta_1), \quad X_2 \sim G(\alpha_2, \beta_2).$$

Proposition 2. Let $(X_1, X_2)^\top \sim BG(\alpha_1, \alpha_2, \beta_1, \beta_2)$. For integers $m \geq 1$ and $n \geq 1$ the product moment is given by

$$E(X_1^m X_2^n) = \frac{\beta_1^m \beta_2^n \Gamma(\alpha_1 + \alpha_2 + m) \Gamma(\alpha_1 + \alpha_2 + n) B(\alpha_1 + m + n, \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2) \Gamma(\alpha_1 + \alpha_2)}, \quad (4.5)$$

where $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ is the beta function. In particular, the covariance and correlation coefficient are

$$\text{Cov}(X_1, X_2) = \frac{\beta_1 \beta_2 \alpha_1 \alpha_2}{\alpha_1 + \alpha_2 + 1}, \quad \text{Corr}(X_1, X_2) = \sqrt{\frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2 + 1}}, \quad (4.6)$$

and the means and variances are

$$E(X_1) = \alpha_1 \beta_1, \quad \text{Var}(X_1) = \alpha_1 \beta_1^2, \quad E(X_2) = \alpha_2 \beta_2, \quad \text{Var}(X_2) = \alpha_2 \beta_2^2. \quad (4.7)$$

4.3 The G-P Analysis — Independence case

Suppose $X \sim G(\alpha_1, \beta_1)$ and $Y \sim G(\alpha_2, \beta_2)$ are independent gamma random variables. We are interested in the linear combination

$$Z = X + aY + b,$$

for some constant $a \neq 0$ and $b \in \mathbb{R}$, and in particular in deriving the distribution of Z and the probability $P(Z > 0)$. Let f_X and f_Y the probability density functions of X and Y , respectively. The following result provides a general expression for the density function of Z .

Theorem 2. Let $X \sim G(\alpha_1, \beta_1)$ and $Y \sim G(\alpha_2, \beta_2)$ be independent. Define the random variable $Z = X + aY + b$. Then the probability density function of Z is given by

$$f_Z(z) = \begin{cases} \int_{(z-b)/a}^\infty f_X(z-ay-b) f_Y(y) dy, & a < 0, z \leq b, \\ \int_0^\infty f_X(z-ay-b) f_Y(y) dy, & a < 0, z > b, \\ 0, & a > 0, z \leq b, \\ \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y) dy, & a > 0, z > b, \end{cases} \quad (4.8)$$

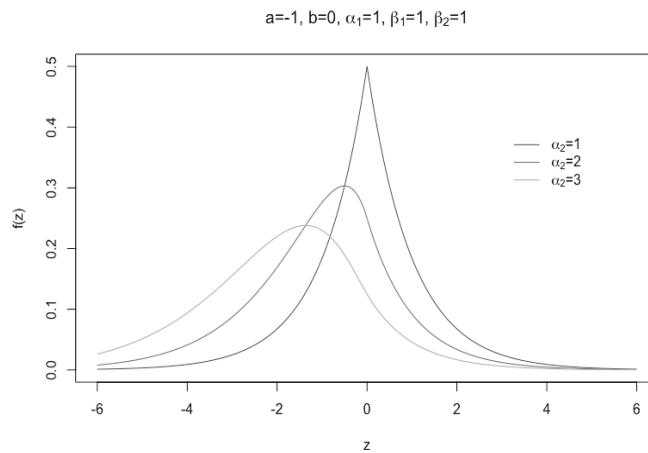
where the integrals are taken over the support $y > 0$.

With this density in hand, one can compute the probability $\mathbb{P}(Z > 0)$ by integrating f_Z over the appropriate region. The resulting expression is given by

$$P(Z > 0) = \begin{cases} \int_0^\infty \int_0^\infty f_X(z-ay-b) f_Y(y) dy dz, & a < 0, b < 0, \\ \left[\int_b^\infty \int_0^\infty + \int_0^b \int_{\frac{z-b}{a}}^\infty \right] f_X(z-ay-b) f_Y(y) dy dz, & a < 0, b \geq 0, \\ \int_0^\infty \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y) dy dz, & a > 0, b < 0, \\ \int_b^\infty \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y) dy dz, & a > 0, b \geq 0. \end{cases} \quad (4.9)$$

Density curves of Z for various parameter combinations, along with their corresponding $P(Z > 0)$ values, are shown in Figure 2.

The probability may be computed directly using the formulas above, and a freely available online calculator is provided at <https://gainprobabilityanalyses.shinyapps.io/independent/>.



Source(s): Figure by the authors

Figure 2. Density of Z with $a=-1$, $b=0$, $\alpha_1=1$, and $\beta_1=\beta_2=1$ for all populations. The shape parameter α_2 takes values 1, 2, and 3, for the first, second and third populations, respectively. The corresponding probabilities $P(Z>0)$ are 0.5000, 0.2500, and 0.1250.

4.4 G-P Analysis — Bivariate Case

Now suppose $\mathbf{X} = (X_1, X_2)^T \sim BG(\alpha_1, \alpha_2, \beta_1, \beta_2)$ is a bivariate gamma random vector as defined earlier, with joint density $g(x_1, x_2)$. Consider the linear combination

$$U = X_1 + aX_2 + b,$$

with $a \neq 0$ and $b \in \mathbb{R}$. The following theorem describes the density function of U .

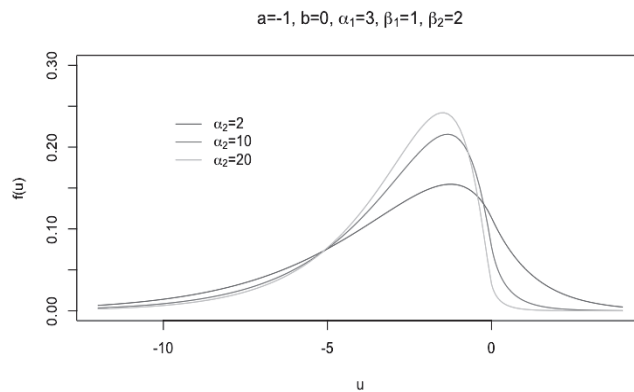
Theorem 3. Assume $(X_1, X_2)^T \sim BG(\alpha_1, \alpha_2, \beta_1, \beta_2)$ and let $U = X_1 + aX_2 + b$. Then the probability density function of U is given by

$$f_U(u) = \begin{cases} \int_{\frac{u-b}{a}}^{\infty} g(u - ax_2 - b, x_2) dx_2, & a < 0, u \leq b, \\ \int_0^{\infty} g(u - ax_2 - b, x_2) dx_2, & a < 0, u > b, \\ 0, & a > 0, u \leq b, \\ \int_0^{\frac{u-b}{a}} g(u - ax_2 - b, x_2) dx_2, & a > 0, u > b. \end{cases} \quad (4.10)$$

The corresponding probability $P(U > 0)$ is

$$P(U > 0) = \begin{cases} \int_0^{\infty} \int_0^{\infty} g(u - ax_2 - b, x_2) dx_2 du, & a < 0, b < 0, \\ \left[\int_b^{\infty} \int_0^{\infty} + \int_0^b \int_{\frac{u-b}{a}}^{\infty} \right] g(u - ax_2 - b, x_2) dx_2 du, & a < 0, b \geq 0, \\ \int_0^{\infty} \int_0^{\frac{u-b}{a}} g(u - ax_2 - b, x_2) dx_2 du, & a > 0, b < 0, \\ \int_b^{\infty} \int_0^{\frac{u-b}{a}} g(u - ax_2 - b, x_2) dx_2 du, & a > 0, b \geq 0. \end{cases} \quad (4.11)$$

The density curves of U for various parameter values, along with their corresponding probabilities $P(U > 0)$, are presented in Figure 3.



Source(s): Figure by the authors

Figure 3. Density of U with $a=-1, b=0, \alpha_1=3, \beta_1=1$, and $\beta_2=2$ for all populations. The shape parameter α_2 takes values 2, 10, and 20 for the first, second, and third populations, respectively. The corresponding probabilities $P(U>0)$ are 0.1448, 0.0415, and 0.0103.

4.5 Monte Carlo Simulation

In this section, we present a Monte Carlo simulation study conducted in R. For the independent case, X and Y were generated from gamma distributions using the `rgamma` function in the ‘stats’ package. In the bivariate case, V_1 and V_2 were also sampled from gamma distributions while W was sampled from a beta distribution using the `rbeta` function. A total of 1,000 simulations were performed to estimate the MLEs and the probabilities $P(Z>0)$ and $P(U>0)$ for various sample sizes $n = 50, 100, 300$, where $Z = X - Y$ and $U = X_1 - X_2 + 0.15$.

The estimated probabilities $P(Z>0)$ for the parameter setting $\alpha_1=0.7, \alpha_2=0.7, \beta_1=0.5, \beta_2=0.5$ are shown in Tables 3. Similarly, the estimated probabilities $P(U>0)$ for $\alpha_1=5, \alpha_2=10, \beta_1=0.7, \beta_2=0.5$ are presented in Tables 4. Absolute biases (|Bias|) and standard errors (SE) of the estimators are likewise reported in the tables. Generally speaking, the MLEs are close to the true parameter values, with bias decreasing as the sample size increases.

Table 3. Simulation results for $P(Z>0)$ in the independent case ($a=-1, b=0$)

$n=50$	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	0.7000	0.7000	0.5000	0.5000	0.5000
MLE	0.7336	0.7329	0.4948	0.4972	0.4993
Bias	0.0336	0.0329	0.0052	0.0028	0.0007
SE	0.1379	0.1318	0.1201	0.1169	
$n=100$	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	0.7000	0.7000	0.5000	0.5000	0.5000
MLE	0.7179	0.7196	0.4961	0.4928	0.5006
Bias	0.0179	0.0196	0.0039	0.0072	0.0006
SE	0.0874	0.0874	0.085	0.0832	
$n=300$	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	0.7000	0.7000	0.5000	0.5000	0.5000
MLE	0.7041	0.7046	0.4973	0.4980	0.4995
Bias	0.0041	0.0046	0.0027	0.0020	0.0005
SE	0.0500	0.0508	0.0501	0.0488	

Source(s): Table by the authors

4.6 Applications in Real Data

The dataset comprises the average length-of-stay (LOS) for patients admitted for acute care in

Table 4. Simulation results for $P(U>0)$ in the dependent case ($a=-1, b=0$)

n=50	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	5	10	0.7	0.5	0.8590
MLE	5.3373	11.2476	0.6831	0.4895	0.8674
Bias	0.3373	1.2476	0.0169	0.0105	0.0084
SE	0.6799	1.2025	0.1412	0.1021	
n=100	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	5	10	0.7	0.5	0.8590
MLE	5.1303	10.7515	0.6957	0.4961	0.8661
Bias	0.1303	0.7515	0.0043	0.0039	0.0071
SE	0.4230	0.8157	0.1018	0.0712	
n=300	α_1	α_2	β_1	β_2	$P(Z>0)$
True Value	5	10	0.7	0.5	0.8590
MLE	5.0481	10.1984	0.6980	0.4980	0.8614
Bias	0.0481	0.1984	0.002	0.002	0.0024
SE	0.2435	0.4187	0.0582	0.0412	

Source(s): Table by the authors

cross hospitals in 10 U.S. states in 1997, available at <https://lib.stat.cmu.edu/data-expo/1997/>. In this example, we compare the average LOS for patients with nervous system conditions (X_1) to those with endocrine system conditions (X_2) within the same hospitals.

The bivariate gamma distribution is adopted to model $(X_1, X_2)'$ because the average LOS cross different departments are likely to be dependent. This dependence may arise from shared institutional factors such as hospital resources and staff expertise, which can simultaneously influence both patient groups. Hospitals with missing data for either variable were excluded from the analysis, resulting in a final sample size of $n = 1079$.

Table 5 summarizes X_1 and X_2 , reporting their means, variances, covariance, correlation, and skewness. MLEs of the bivariate gamma parameters, computed numerically in R, are $\hat{\alpha}_1 = 8.6010$, $\hat{\alpha}_2 = 63.8334$, $\hat{\beta}_1 = 0.7116$, and $\hat{\beta}_2 = 0.6898$. Figure 4 displays histograms of X_1 and X_2 , respectively, overlaid with their fitted gamma density curves $G(\hat{\alpha}_i, \hat{\beta}_i)$, $i = 1, 2$.

Table 5. Summary statistics for the average LOS of patients in the nervous system (X_1) and endocrine (X_2) departments

	Mean	Std. Dev.	Cov	Cor	Skewness
X_1	6.36	2.19	4.00	0.87	0.79
X_2	5.70	2.11			1.06

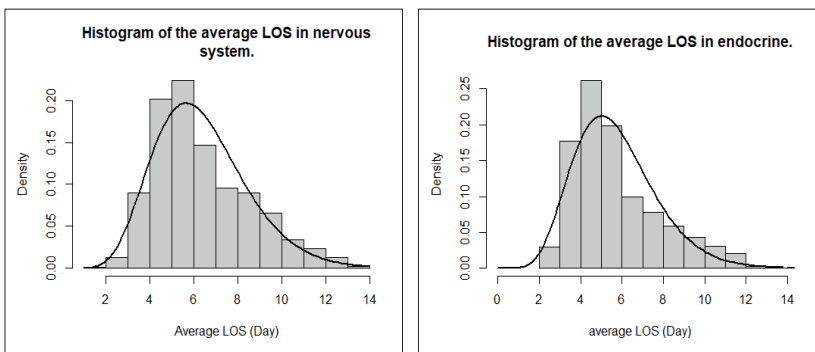
Source(s): Table by the authors

Building on the analysis of the LOS data, in the context of matched paired data, the G-P analysis using the bivariate gamma distribution provides an intuitive probabilistic interpretation of the proportion of acute care cases in which the LOS for patients with nervous system conditions exceeds that for those with endocrine system conditions. Specially,

$$P(X_1 > X_2) = 0.5739.$$

5. Comparative Framework

To clarify the conceptual distinctions among traditional NHST, the APP and G-P analysis, Table 6 provides a comparative summary across several important dimensions. The comparison highlights differences in inferential objectives, interpretation, timing of analysis, dependence on sample size, and underlying statistical philosophy. In particular, while NHST focuses on hypothesis rejection through (p)-values, APP emphasizes estimation reliability before data collection, and G-P analysis provides an estimation-based probabilistic interpretation after estimation. This



Source(s): Figure by the authors
Figure 4. Histograms and fitted gamma distributions for the average LOS data across two medical conditions

Table 6. Comparison of NHST, A Priori Procedure (APP), and Gain–Probability (G–P) Analysis

Dimension	NHST	A Priori Procedure (APP)	Gain–Probability (G–P)
Primary Goal	Reject H_0 based on observed data.	Ensure estimation reliability before data collection.	Quantify probability that one group outperforms another.
Inferential Quantity	$p = \Pr(\text{data} H_0)$	$\Pr(\hat{\theta} - \theta < \delta) \geq 1 - \alpha$	$G = \Pr(X > Y)$
Interpretive Focus	Binary decision.	Precision and design adequacy.	Probabilistic meaning.
Timing	Post hoc.	A priori.	Post-estimation.
Dependence on n	High; inflates trivial effects.	Explicitly controlled.	Interpretable irrespective of n.
Philosophy	Model rejection.	Design-based reliability.	Estimation-based comparison.
Key References	Fisher (1925); Neyman and Pearson (1933)	Trafimow et al. (2024)	Tong et al. (2022b); Trafimow et al. (2024)

Source(s): Table by the authors

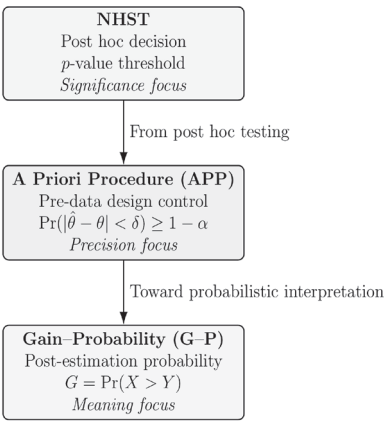
comparison further illustrates the broader transition from dichotomous significance testing toward estimation-oriented and decision-relevant statistical inference.

Figure 5 illustrates the conceptual evolution of modern statistical inference from traditional NHST toward estimation-oriented and probabilistic frameworks. The progression emphasizes a shift away from dichotomous post hoc decision-making based on p-value thresholds toward approaches that prioritize estimation reliability and interpretable probabilistic conclusions. In this framework, the APP serves as a bridge by focusing on pre-data precision control, while the G–P analysis extends the inferential process to meaningful post-estimation probability comparisons. Together, these approaches reflect an increasing emphasis on estimation quality, interpretability, and practical decision relevance in statistical analysis.

6. Conclusion and Future Directions

This work has examined and contrasted three inferential paradigms–Null Hypothesis Significance Testing (NHST), the A Priori Procedure (APP), and Gain–Probability (G–P) analysis–within the broader context of statistical reform and the reproducibility movement. While NHST remains a dominant framework, its reliance on p-values and binary decision rules has led to substantial criticism, particularly regarding interpretability, sensitivity to sample size, and the gap between statistical and practical significance.

In contrast, the APP and G–P frameworks offer estimation- and probability-based alternatives that address many of these shortcomings. The APP promotes pre-data planning and estimation reliability by guiding researchers toward appropriate sample sizes based on desired levels of precision. The G–P framework, in turn, quantifies the probability that one outcome exceeds another and by varying degrees offering a continuous and interpretable measure of effect dominance that



Source(s): Figure by the authors
Figure 5. Conceptual evolution from dichotomous testing (NHST) to estimation-based and probabilistic inference (APP and G–P)

aligns more closely with scientific reasoning and decision-making under uncertainty.

Methodologically, we demonstrated the application of APP for estimating parameters in linear regression models with skew-normal errors, supported by theoretical results, simulation studies, and real-data illustrations. We also investigated G–P analysis for gamma-distributed outcomes, both in independent and dependent settings, again supported by simulation evidence and empirical applications.

Collectively, these frameworks provide a rigorous and interpretable alternative to classical significance testing. By prioritizing estimation accuracy, probabilistic reasoning, and practical understanding of effects, they offer a unified foundation for transparent and reproducible inference. As statistical methodology continues to evolve, the integration of APP and G–P analysis into standard practice holds promise for advancing both theoretical rigor and applied insight in scientific research.

Several promising avenues for future research emerge from this comparative and applied study. First, further theoretical development is needed to extend the APP and G–P frameworks to a broader class of statistical models, including generalized linear models, mixed-effects models, and nonparametric settings. Such extensions would increase the accessibility and utility of estimation-based inference in diverse applied domains.

Second, additional work is warranted on refining computational tools and user-friendly software implementations of that APP and G–P analysis for practitioners. In particular, interactive applications and R/Python libraries could facilitate adoption in educational, clinical, and policy contexts.

Third, empirical investigations across disciplines—especially in psychology, medicine, and economics—could assess how APP- and G–P-based inference influences replicability, effect interpretation, and decision quality in practice. Comparing these approaches to traditional NHST in real-world replication studies may further illuminate their benefits and limitations.

Finally, from a pedagogical perspective, incorporating estimation- and probability-based inference into undergraduate and graduate curricula would help cultivate a new generation of researchers attuned to transparency, uncertainty quantification, and practical significance. This cultural shift toward estimation-based thinking is essential for the long-term sustainability of credible scientific inference.

References

- Amrhein, V., Greenland, S., and McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748):305–307.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12):997–1003.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1):7–29.
- Fisher, R. A. (1925). Theory of statistical estimation. In *Mathematical proceedings of the Cambridge philosophical society*, volume 22, pages 700–725. Cambridge University Press.
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33(5):587–606.
- Goodman, S. N. (2008). A dirty dozen: twelve p-value misconceptions. *Seminars in Hematology*, 45(3):135–140.
- Hubbard, R. and Bayarri, M. J. (2003). To p or not to p? *Psychological Bulletin*, 133(4):605–617.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. (2019). *Continuous multivariate distributions, Volume 1: Models and applications*, volume 1. John Wiley & sons.
- Ma, Z., Wang, T., Wei, Z., and Zhu, X. (2022). Inferences on location parameters based on independent multivariate skew normal distributions. *Asian Journal of Economics and Banking*, 6(2):270–281.
- Nadarajah, S. and Gupta, A. K. (2006). Some bivariate gamma distributions. *Applied mathematics letters*, 19(8):767–774.
- Neyman, J. and Pearson, E. S. (1933). IX. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337.
- Tong, T., Trafimow, D., Wang, T., Wang, C., Hu, L., and Chen, X. (2022a). The a priori procedure (app) for estimating regression coefficients in linear models. *Methodology*, 18(3):203–220.
- Tong, T., Wang, T., Trafimow, D., and Wang, C. (2022b). The probability of being better or worse off, and by how much, depending on experimental conditions with skew normal populations. In *International Conference of the Thailand Econometrics Society*, pages 261–284. Springer.
- Trafimow, D. and Marks, M. (2015). Editorial in basic and applied social psychology. *Basic and Applied Social Psychology*, 37:1–2.
- Trafimow, D., Roth, N., Xu, L., Toomasian, D., Perrello, A., Tong, T., Wang, T., Choy, S. B., Chen, X., Wang, C., et al. (2023). Surprising implications of differences in locations versus differences in means. *Methodology*, 19(2):152–169.
- Trafimow, D., Tong, T., Wang, T., Choy, S. T. B., Hu, L., Chen, X., Wang, C., and Wang, Z. (2024). Improving inferential analyses: Predata and postdata. *Psychological Methods*. Advance online publication.
- Trafimow, D., Wang, T., and Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! *Educational and Psychological Measurement*, 79(1):129–150.
- Wasserstein, R. L. and Lazar, N. A. (2016). The asa’s statement on p-values: context, process, and purpose. *The American Statistician*, 70(2):129–133.
- Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019). *Moving to a world beyond “ $p < 0.05$ ”*.

Corresponding author

Tonghui Wang can be contacted at: twang@nmsu.edu